

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
REQUEST FOR FILING APPLICATION UNDER RULE 1.53(b)**

jc613 U.S. PTO  
09/705911  
11/06/00

Pursuant to 37 CFR 1.53(b), please file a ☐ continuation/☒ divisional of the pending prior PATENT APPLICATION of:

Inventor: HERMON-TAYLOR et al.

Serial No. 09/091,538

Filed: September 16, 1998

For: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY

Assistant Commissioner for Patents

Washington, DC 20231

Sir:

Atty Dkt.: 117-323

C# M#

Date: November 6, 2000

Group: 1645

Examiner: R. Baskar

This request for filing under Rule 53(b) is made by the following named inventor(s) (using the above-identified title):  
Inventor(s): HERMON-TAYLOR et al.

- ☒ Attached is a true copy of the prior application as originally filed including the specification, claims, Oath/Declaration and drawings (if any) and abstract (if any). No amendments (if any) referenced in the Oath or Declaration filed to complete the prior application introduced new matter.
- ☒ Priority is hereby claimed under 35 USC 119 based on the following foreign applications, the entire content of which is hereby incorporated by reference in this application:

Application Number	Country	Day/Month/Year/Filed
9526178.0	Great Britain	21 December 1995
PCT/GB96/03221	PCT	23 December 1996

☒ certified copy(ies) of foreign application(s) attached or

☐ already filed on \_\_\_\_\_ in prior appln. no. \_\_\_\_\_

☒ already filed in 09/091,538 filed September 16, 1998

Please amend the specification by inserting before the first line: -- This application claims the benefit of U.S.

Provisional Application No. \_\_\_\_\_, filed \_\_\_\_\_, the entire content of which is hereby incorporated by reference in this application.--

☒ The prior application is assigned to St. George's Hospital Medical School.

☒ Power of Attorney has been granted to B.J. Sadoff et al, Reg. No. 36,663 of Nixon & Vanderhye P.C., 1100 N. Glebe Rd., 8<sup>th</sup> Floor, Arlington, VA 22201.

☒ Address all future communications to: Nixon & Vanderhye P.C., 1100 N. Glebe Rd., 8<sup>th</sup> Floor, Arlington, VA 22201.

☒ Please amend the specification by inserting before the first line --This is a divisional of application Serial No. 09/091,538, filed September 16, 1998, now pending, which is a 371 application of PCT/GB96/03221, filed December 23, 1996 the entire content of which is hereby incorporated by reference in this application.--

☐ "Small entity" statement of record. ☐ "Small entity" statement attached.

☐ Petition filed in prior application to extend its life to insure co-pendency.

☒ The Examiner's attention is directed to the prior art cited in the parent application by applicant and/or Examiner for the reasons stated therein.

☒ Please enter the attached and/or below preliminary amendment prior to calculation of filing fee:

- ☒ The entire disclosure of the prior application above-referenced is considered as being part of the disclosure of this new application and is hereby incorporated by reference therein.

**FILING FEE IS BASED ON CLAIMS AS FILED LESS ANY HEREWITH CANCELED**

Basic Filing Fee		\$	710.00
Total effective claims	23 - 20 (at least 20) = 3	x \$ 18.00	\$ 54.00
Independent claims	4 - 3 (at least 3) = 1	x \$ 80.00	\$ 80.00
If any proper multiple dependent claims now added for first time, add \$270.00 (ignore improper)		\$	270.00
		<b>SUBTOTAL</b>	\$ 1,114.00
If "small entity," then enter half (1/2) of subtotal and subtract		-(	0.00)
		<b>SECOND SUBTOTAL</b>	\$ 1,114.00
Assignment Recording Fee (\$40.00)		\$	0.00
		<b>TOTAL FEE ENCLOSED</b>	\$ 1,114.00

Any future submission requiring an extension of time is hereby stated to include a petition for such time extension.

The Commissioner is hereby authorized to charge any deficiency in the fee(s) filed, or asserted to be filed, or which should have been filed herewith (or with any paper hereafter filed in this application by this firm) to our **Account No. 14-1140**. A duplicate copy of this sheet is attached.

1100 North Glebe Road, 8<sup>th</sup> Floor  
Arlington, Virginia 22201-4714  
Telephone: (703) 816-4000  
Facsimile: (703) 816-4100  
BJS:jls

**NIXON & VANDERHYE P.C.**

By Atty: B.J. Sadoff, Reg. No. 36,663

Signature: \_\_\_\_\_

472678

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

HERMON-TAYLOR et al.

Atty. Ref.: 117-323

Divisional of Serial No. 09/091,538

Group: Unassigned

Filed: Herewith

Examiner: Unassigned

For: NOVEL POLYNUCLEOTIDES AND  
POLYPEPTIDES IN PATHOGENIC  
MYCOBACTERIA AND THEIR USE AS  
DIAGNOSTICS, VACCINES AND TARGETS  
FOR CHEMOTHERAPY

\* \* \* \* \*

November 6, 2000

Assistant Commissioner for Patents  
Washington, DC 20231

**PRELIMINARY AMENDMENT**

Sir:

Entry and consideration of the following amendments and remarks are requested.

**IN THE SPECIFICATION:**

Amend the specification as follows.

Insert the attached Sequence Listing after the claims pages.

**IN THE CLAIMS:**

Amend the claims as follows.

Cancel claims 2, 3, 16 and 17, without prejudice.

4. (Amended) A polynucleotide in substantially isolated form which encodes a polypeptide according to claim 1 [any one of claims 1 to 3].

8. (Amended) A polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide as defined in claim 5 [any one of claims 4 to 7], optionally carrying a revealing label.

9. (Amended) A recombinant vector carrying a polynucleotide as defined in claim 5 [any one of claims 4 to 7].

10. (Amended) An antibody capable of binding a polypeptide or fragment thereof as defined in claim 1 [any one of claims 1 to 3].

12. (Amended) A test kit for detecting the presence or absence of a pathogenic mycobacterium in a sample which comprises a polynucleotide according to claim 4 [any one of claims 4 to 8], a polypeptide according to claim 1 [any one of claims 1 to 3], a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, or an antibody according to claim 10 [, any one of claims 10 or 11].

13. (Amended) A method of detecting the presence or absence of antibodies in an animal or human, against a pathogenic mycobacteria in a sample which comprises:

(a) providing a polypeptide according to [any one of claims 1 to 3] claim 1 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which comprises an epitope;

(b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody—antigen complex; and

(c) determining whether antibody-antigen complex comprising said polypeptide is formed.

14. (Amended) A method of detecting the presence or absence of a polypeptide according to [any one of claims 1 to 3] claim 1 or a polypeptide which comprises a sequence

selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto in a biological sample which method which comprises:

- (a) providing an antibody according to claim 10 [any one of claims 10 and 11];
- (b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said antibody is formed.

15. (Amended) A method of detecting the presence or absence of cell mediated immune reactivity in an animal or human, to a polypeptide according to claim 1 [claims 1 to 3] or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which method comprises

- (a) providing a polypeptide according to claim 1 [any one of claims 1 to 3] or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which comprises an epitope;
- (b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator or reaction to occur; and
- (c) detecting the presence of said cytokine or mediator or cellular response in the incubate.

18. (Amended) A method of treating or preventing mycobacterial disease in an animal or human caused by mycobacteria which express a polypeptide according to [claims 1 to 3] claim 1 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which method comprises vaccinating or treating an animal or human with an effective amount of said polypeptide.

19. (Amended) A method of treating or preventing mycobacterial diseases in animals or humans caused by mycobacteria containing the polynucleotide of Seq.ID.No: 3 or 4, which method comprises vaccinating or treating an animal or human with an effective amount of a polynucleotide according to claim 4 [claims 4 to 7], a vector according to claim 9 or a polynucleotide which encodes a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto.

20. (Amended) A method according to claim 18 [claims 18 or 19] for increasing the in vivo susceptibility of mycobacteria to antimicrobial drugs.

21. (Amended) A normally pathogenic mycobacterium, whose pathogenicity is mediated in all or in part by the presence or the expression of a polypeptide as defined in [any one of claims 1 to 3] claim 1 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which mycobacterium harbours an attenuating mutation in a gene encoding one of the said polypeptides.

#### **REMARKS**

The claims have been amended to reduce the filing fees and delete improper multiple dependencies.

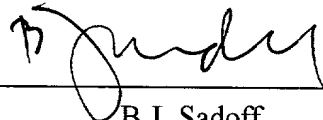
The specification has been amended to include a Sequence Listing, a copy of which was filed in the parent Application No.09/091,538. The attached paper copy of the Sequence Listing is the same as the paper and computer readable copies of the Sequence Listing submitted in Application No. 09/091,538. The Office is requested to use the computer readable form of the Sequence Listing in the parent Application No. 09/091,538, in the present application. A separate Request to this effect is attached. No new matter has been added.

An early and favorable Action on the merits is requested.

Respectfully submitted,

**NIXON & VANDERHYE P.C.**

By: \_\_\_\_\_



B.J. Sadoff  
Reg. No. 36,663

BJS:rdw  
1100 North Glebe Road, 8th Floor  
Arlington, VA 22201-4714  
Telephone: (703) 816-4091  
Facsimile: (703) 816-4100

Novel polynucleotides and polypeptides in pathogenic mycobacteria and their use as diagnostics, vaccines and targets for chemotherapy.

This invention relates to the novel polynucleotide sequence we  
5 have designated "GS" which we have identified in pathogenic  
mycobacteria. GS is a pathogenicity island within 8kb of DNA  
comprising a core region of 5.75kb and an adjacent transmissible  
element within 2.25kb. GS is contained within *Mycobacterium*  
10 *paratuberculosis*, *Mycobacterium avium* subsp. *silvaticum* and some  
pathogenic isolates of *M. avium*. Functional portions of the core  
region of GS are also represented by regions with a high degree  
of homology that we have identified in cosmids containing genomic  
DNA from *Mycobacterium tuberculosis*.

15 Background to the invention

*Mycobacterium tuberculosis* (Mtb) is a major cause of global  
diseases of humans as well as animals. Although conventional  
methods of diagnosis including microscopy, culture and skin  
testing exist for the recognition of these diseases, improved  
20 methods particularly new immunodiagnosics and DNA-based  
detection systems are needed. Drugs used to treat tuberculosis  
are increasingly encountering the problem of resistant organisms.  
New drugs targeted at specific pathogenicity determinants as well  
as new vaccines for the prevention and treatment of tuberculosis  
25 are required. The importance of Mtb as a global pathogen is  
reflected in the commitment being made to sequencing the entire  
genome of this organism. This has generated a large amount of  
DNA sequence data of genomic DNA within cosmid and other  
libraries. Although the DNA sequence is known in the art, the  
30 functions of the vast majority of these sequences, the proteins  
they encode, the biological significance of these proteins, and  
the overall relevance and use of these genes and their products  
as diagnostics, vaccines and targets for chemotherapy for  
tuberculous disease, remains entirely unknown.

35 *Mycobacterium avium* subsp. *silvaticum* (Mavs) is a pathogenic  
mycobacterium causing diseases of animals and birds, but it can

also affect humans. *Mycobacterium paratuberculosis* (*Mptb*) causes chronic inflammation of the intestine in many species of animals including primates and can also cause Crohn's disease in humans. *Mptb* is associated with other chronic inflammatory diseases of humans such as sarcoidosis. Subclinical *Mptb* infection is widespread in domestic livestock and is present in milk from infected animals. The organism is more resistant to pasteurisation than *Mtb* and can be conveyed to humans in retail milk supplies. *Mptb* is also present in water supplies, particularly those contaminated with run-off from heavily grazed pastures. *Mptb* and *Mavs* contain the insertion elements IS900 and IS902 respectively, and these are linked to pathogenicity in these organisms. IS900 and IS902 provide convenient highly specific multi-copy DNA targets for the sensitive detection of these organisms using DNA-based methods and for the diagnosis of infections in animals and humans. Much improvement is however required in the immunodiagnosis of *Mptb* and *Mavs* infections in animals and humans. *Mptb* and *Mavs* are in general, resistant in vivo to standard anti-tuberculous drugs. Although substantial clinical improvements in infections caused by *Mptb*, such as Crohn's disease, may result from treatment of patients with combinations of existing drugs such as Rifabutin, Clarithromycin or Azithromycin, additional effective drug treatments are required. Furthermore, there is an urgent need for effective vaccines for the prevention and treatment of *Mptb* and *Mavs* infections in animals and humans based upon the recognition of specific pathogenicity determinants.

Pathogenicity islands are, in general, 7-9kb regions of DNA comprising a core domain with multiple ORFs and an adjacent transmissible element. The transmissible element also encodes proteins which may be linked to pathogenicity, such as by providing receptors for cellular recognition. Pathogenicity islands are envisaged as mobile packages of DNA which, when they enter an organism, assist in bringing about its conversion from a non-disease-causing to a disease-causing strain.

#### Description of the Drawings



Figure 1(a) and (b) shows a linear map of the pathogenicity island GS in *Mavs* (Fig 1a) and in *Mptb* (Fig 1b). The main open reading frames are illustrated as ORFs A to H. ORFs A to F are found within the core region of GS. ORFs G and H are encoded by the adjacent transmissible element portion of GS.

#### Disclosure of the invention

Using a DNA-based differential analysis technology we have discovered and characterised a novel polynucleotide in *Mptb* (isolates 0022 from a Guernsey cow and 0021 from a red deer). This polynucleotide comprises the gene region we have designated GS. GS is found in *Mptb* using the identifier DNA sequences Seq.ID.No 1 and 2 where the Seq.ID No2 is the complementary sequence of Seq.ID No 1. GS is also identified in *Mavs*. The complete DNA sequence incorporating the positive strand of GS from an isolate of *Mavs* comprising 7995 nucleotides, including the core region of GS and adjacent transmissible element, is given in Seq.ID No.3. DNA sequence comprising 4435 bp of the positive strand of GS obtained from an isolate of *Mptb* including the core region of GS (nucleotides 1614 to 6047 of GS in *Mavs*) is given in Seq.ID No 4. The DNA sequence of GS from *Mptb* is highly (99.4%) homologous to GS in *Mavs*. The remaining portion of the DNA sequence of GS in *Mptb*, is readily obtainable by a person skilled in the art using standard laboratory procedures. The entire functional DNA sequence including core region and transmissible element of GS in *Mptb* and *Mavs* as described above, comprise the polynucleotide sequences of the invention.

There are 8 open reading frames (ORFs) in GS. Six of these designated GSA, GSB, GSC, GSD, GSE and GSF are encoded by the core DNA region of GS which, characteristically for a pathogenicity island, has a different GC content than the rest of the microbial genome. Two ORFs designated GSG and GSH are encoded by the transmissible element of GS whose GC content resembles that of the rest of the mycobacterial genome. The ORF GSH comprises two sub-ORFs H<sub>1</sub> H<sub>2</sub> on the complementary DNA strand linked by a programmed frameshifting site so that a single polypeptide is translated from the ORF GSH. The nucleotide

- 4 -

sequences of the 8 ORFs in GS and their translations are shown in Seq. ID No 5 to Seq.ID No 29 as follows:

- 5 ORF A: Seq. ID No 5 Nucleotides 50 to 427 of GS from *Mavs*  
Seq. ID No 6 Amino acid sequence encoded by Seq.ID No 5.
- ORF B: Seq. ID No 7 Nucleotides 772 to 1605 of GS from *Mavs*  
Seq. ID No 8 Amino acid sequence encoded by Seq.ID No 7.
- 10 ORF C: Seq. ID No 9 Nucleotides 1814 to 2845 of GS from *Mavs*  
Seq. ID No 10 Amino acid sequence encoded by Seq.ID No 9.  
Seq. ID No 11 Nucleotides 201 to 1232 of GS from *Mptb*  
Seq. ID No 12 Amino acid sequence encoded by Seq.ID No 11
- 15 ORF D: Seq. ID No 13 Nucleotides 2785 to 3804 of GS from *Mavs*  
Seq. ID No 14 Amino acid sequence encoded by Seq.ID No 13.  
Seq. ID No 15 Nucleotides 1172 to 2191 of GS from *Mptb*  
Seq. ID No 16 Amino acid sequence encoded by Seq.ID No 15.
- 20
- ORF E: Seq. ID No 17 Nucleotides 4080 to 4802 of GS from *Mavs*  
Seq. ID No 18 Amino acid sequence encoded by Seq.ID No 17.  
Seq. ID No 19 Nucleotides 2467 to 3189 of GS from *Mptb*  
Seq. ID No 20 Amino acid sequence encoded by Seq.ID No 19.
- 25
- ORF F: Seq. ID No 21 Nucleotides 4947 to 5747 of GS from *Mavs*  
Seq. ID No 22 Amino acid sequence encoded by Seq.ID No 21.  
Seq. ID No 23 Nucleotides 3335 to 4135 of GS from *Mptb*  
Seq. ID No 24 Amino acid sequence encoded by Seq.ID No 23.
- 30

- 5 -

ORF G: Seq. ID No 25 Nucleotides 6176 to 7042 of GS from *Mavs*  
Seq. ID No 26 Amino acid sequence encoded by  
Seq.ID No 25.

ORF H: Seq.ID No 27 Nucleotides 7953 to 6215 from *Mavs*.

5 ORF H<sub>1</sub>: Seq.ID No 28 Amino acid sequence encoded by  
nucleotides 7953 to 7006 of Seq.ID No 27

ORF H<sub>2</sub>: Seq.ID No 29 Amino acid sequence encoded by  
nucleotides 7009 to 6215 of Seq.ID No 27

10 The polynucleotides in *Mtb* with homology to the ORFs B, C, E and  
F of GS in *Mptb* and *Mavs*, and the polypeptides they are now known  
to encode as a result of our invention, are as follows:

15 ORF B: Seq.ID No 30 Cosmid MTCY277 nucleotides 35493 to  
34705  
Seq.ID No 31 Amino acid sequence encoded by Seq.ID  
No30.

ORF C: Seq.ID No 32 Cosmid MTCY277 nucleotides 31972 to 32994  
Seq.ID No 33 Amino acid sequence encoded by Seq.ID  
No32.

20 ORF E: Seq.ID No 34 Cosmid MTCY277 nucleotides 34687 to 33956  
Seq.ID No 35 Amino acid sequence encoded by Seq.ID  
No34.

ORF E: Seq.ID No 36 Cosmid MTO24 nucleotides 15934 to 15203  
Seq.ID No 37 Amino acid sequence encoded by Seq.ID  
No36.

25 ORF F: Seq.ID No38 Cosmid MTO24 nucleotides 15133 to 14306  
Seq.ID No 39 Amino acid sequence encoded by Seq.ID  
No38.

The proteins and peptides encoded by the ORFs A to H in *Mptb* and  
*Mavs* and the amino acid sequences from homologous genes we have

- 6 -

discovered in Mtb given in Seq.ID Nos 31, 33, 35, 37 and 39, as described above and fragments thereof, comprise the polypeptides of the invention. The polypeptides of the invention are believed to be associated with specific immunoreactivity and with the pathogenicity of the host micro-organisms from which they were obtained.

The present invention thus provides a polynucleotide in substantially isolated form which is capable of selectively hybridising to sequence ID Nos 3 or 4 or a fragment thereof. The polynucleotide fragment may alternatively comprise a sequence selected from the group of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. The invention further provides a polynucleotide in substantially isolated form whose sequence consists essentially of a sequence selected from the group Seq ID Nos. 30, 32, 34, 36 and 38, or a corresponding sequence selectively hybridizable thereto, or a fragment of said sequence or corresponding sequence.

The invention further provides diagnostic probes such as a probe which comprises a fragment of at least 15 nucleotides of a polynucleotide of the invention, or a peptide nucleic acid or similar synthetic sequence specific ligand, optionally carrying a revealing label. The invention also provides a vector carrying a polynucleotide as defined above, particularly an expression vector.

The invention further provides a polypeptide in substantially isolated form which comprises any one of the sequences selected from the group consisting Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39, or a polypeptide substantially homologous thereto. The invention additionally provides a polypeptide fragment which comprises a fragment of a polypeptide defined above, said fragment comprising at least 10 amino acids and an epitope. The invention also provides polynucleotides in substantially isolated form which encode polypeptides of the invention, and vectors which comprise such polynucleotides, as well as antibodies capable of binding such polypeptides. In an additional aspect, the invention provides

kits comprising polynucleotides, polypeptides, antibodies or synthetic ligands of the invention and methods of using such kits in diagnosing the presence or absence of mycobacteria in a sample. The invention also provides pharmaceutical compositions comprising polynucleotides of the invention, polypeptides of the invention or antisense probes and the use of such compositions in the treatment or prevention of diseases caused by mycobacteria. The invention also provides polynucleotide prevention and treatment of infections due to GS-containing pathogenic mycobacteria in animals and humans and as a means of enhancing in vivo susceptibility of said mycobacteria to antimicrobial drugs. The invention also provides bacteria or viruses transformed with polynucleotides of the invention for use as vaccines. The invention further provides *Mptb* or *Mavs* in which all or part of the polynucleotides of the invention have been deleted or disabled to provide mutated organisms of lower pathogenicity for use as vaccines in animals and humans. The invention further provides *Mtb* in which all or part of the polynucleotides encoding polypeptides of the invention have been deleted or disabled to provide mutated organisms of lower pathogenicity for use as vaccines in animals and humans.

A further aspect of the invention is our discovery of homologies between the ORFs B, C and E in GS on the one hand, and *Mtb* cosmid MTCY277 on the other (data from Genbank database using the computer programmes BLAST and BLIXEM). The homologous ORFs in MTCY277 are adjacent to one another consistent with the form of another pathogenicity island in *Mtb*. A further aspect of the invention is our discovery of homologies between ORFs E and F in GS, and *Mtb* cosmid MTO24 (also Genbank, as above) with the homologous ORFs close to one another. The use of polynucleotides and polypeptides from *Mtb* (Seq. ID Nos 30, 31, 32, 33, 34, 35, 36, 37, 38 and 39) in substantially isolated form as diagnostics, vaccines and targets for chemotherapy, for the management and prevention of *Mtb* infections in humans and animals, and the processes involved in the preparation and use of these diagnostics, vaccines and new chemotherapeutic agents, comprise further aspects of the invention.

Detailed description of the invention.A. Polynucleotides

Polynucleotides of the invention as defined herein may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides or peptide nucleic acids. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to couple the said polynucleotide to a solid phase or to enhance the recognition, the *in vivo* activity, or the lifespan of polynucleotides of the invention.

A number of different types of polynucleotides of the invention are envisaged. In the broadest aspect, polynucleotides and fragments thereof capable of hybridizing to SEQ ID NO:3 or 4 form a first aspect of the invention. This includes the polynucleotide of SEQ ID NO: 3 or 4. Within this class of polynucleotides various sub-classes of polynucleotides are of particular interest.

One sub-class of polynucleotides which is of interest is the class of polynucleotides encoding the open reading frames A, B, C, D, E, F, G and H, including SEQ ID NOs:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. As discussed below, polynucleotides encoding ORF H include the polynucleotide sequences 7953 to 7006 and 7009 to 6215 within SEQ ID NO: 27, as well as modified sequences in which the frame-shift has been modified so that the two sub-reading frames are placed in a single reading frame. This may be desirable where the polypeptide is to be produced in recombinant expression systems.

The invention thus provides a polynucleotide in substantially isolated form which encodes any one of these ORFs or combinations

- 9 -

thereof. Combinations thereof includes combinations of 2, 3, 4, 5 or all of the ORFs. Polynucleotides may be provided which comprise an individual ORF carried in a recombinant vector including the vectors described herein. Thus in one preferred aspect the invention provides a polynucleotide in substantially isolated form capable of selectively hybridizing to the nucleic acid comprising ORFs A to F of the core region of the *Mptb* and *Mavs* pathogenicity islands of the invention. Fragments thereof corresponding to ORFs A to E, B to F, A to D, B to E, A to C, B to D or any two adjacent ORFs are also included in the invention.

Polynucleotides of the invention will be capable of selectively hybridizing to the corresponding portion of the GS region, or to the corresponding ORFs of *Mtb* described herein. The term "selectively hybridizing" indicates that the polynucleotides will hybridize, under conditions of medium to high stringency (for example 0.03 M sodium chloride and 0.03 M sodium citrate at from about 50°C to about 60°C) to the corresponding portion of SEQ ID NO:3 or 4 or the complementary strands thereof but not to genomic DNA from mycobacteria which are usually non-pathogenic including non-pathogenic species of *M.avium*. Such polynucleotides will generally be generally at least 68%, e.g. at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the corresponding DNA of GS. The corresponding portion will be of over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

By "corresponding portion" it is meant a sequence from the GS region of the same or substantially similar size which has been determined, for example by computer alignment, to have the greatest degree of homology to the polynucleotide.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably 30 nucleotides forms one aspect of the invention, as

does a polynucleotide which is at least 90% homologous over 40 nucleotides.

A further class of polynucleotides of the invention is the class of polynucleotides encoding polypeptides of the invention, the polypeptides of the invention being defined in section B below. Due to the redundancy of the genetic code as such, polynucleotides may be of a lower degree of homology than required for selective hybridization to the GS region. However, when such polynucleotides encode polypeptides of the invention these polynucleotides form a further aspect. It may for example be desirable where polypeptides of the invention are produced recombinantly to increase the GC content of such polynucleotides. This increase in GC content may result in higher levels of expression via codon usage more appropriate to the host cell in which recombinant expression is taking place.

An additional class of polynucleotides of the invention are those obtainable from cosmids MTCY277 and MT024 (containing *Mtb* genomic sequences), which polynucleotides consist essentially of the fragment of the cosmid containing an open reading frame encoding any one of the homologous ORFs B, C, E or F respectively. Such polynucleotides are referred to below as *Mtb* polynucleotides. However, where reference is made to polynucleotides in general such reference includes *Mtb* polynucleotides unless the context is explicitly to the contrary. In addition, the invention provides polynucleotides which encode the same polypeptide as the abovementioned ORFs of *Mtb* but which, due to the redundancy of the genetic code, have different nucleotide sequences. These form further *Mtb* polynucleotides of the invention. Fragments of *Mtb* polynucleotides suitable for use as probes or primers also form a further aspect of the invention.

The invention further provides polynucleotides in substantially isolated form capable of selectively hybridizing (where selectively hybridizing is as defined above) to the *Mtb* polynucleotides of the invention.



- 11 -

The invention further provides the *Mtb* polynucleotides of the invention linked, at either the 5' and/or 3' end to polynucleotide sequences to which they are not naturally contiguous. Such sequences will typically be sequences found in cloning or expression vectors, such as promoters, 5' untranslated sequence, 3' untranslated sequence or termination sequences. The sequences may also include further coding sequences such as signal sequences used in recombinant production of proteins.

Further polynucleotides of the invention are illustrated in the accompanying examples.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels or a probe linked covalently to a solid phase, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 or more nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Primers of the invention which are preferred include primers directed to any part of the ORFs defined herein. The ORFs from other isolates of pathogenic mycobacteria which contain a GS region may be determined and conserved regions within each individual ORF may be identified. Primers directed to such conserved regions form a further preferred aspect of the invention. In addition, the primers and other polynucleotides of the invention may be used to identify, obtain and isolate ORFs capable of selectively hybridizing to the polynucleotides of the invention which are present in pathogenic mycobacteria but which are not part of a pathogenicity island in that particular species of bacteria. Thus in addition to the ORFs B, C, E and F which have been identified in *Mtb*, similar ORFs may be identified in other pathogens and ORFs corresponding to the GS ORFs C, D, E, F and H, may also be identified.

Polynucleotides such as DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

- 5 In general, primers will be produced by synthetic means, involving a step-wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art. Longer polynucleotides will generally be produced using  
10 recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair or primers (e.g. of about 15-30 nucleotides) to a region of GS, which it is desired to clone, bringing the primers into contact with genomic DNA from a mycobacterium or a vector carrying the  
15 GS sequence, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme  
20 recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the GS or ORF sequences described herein, as well as further genomic clones containing full open reading frames. Although in general such  
25 techniques are well known in the art, reference may be made in particular to Sambrook J., Fritsch EF., Maniatis T (1989). Molecular cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.

- Polynucleotides which are not 100% homologous to the sequences  
30 of the present invention but fall within the scope of the invention can be obtained in a number of ways.

Other isolates or strains of pathogenic mycobacteria will be expected to contain allelic variants of the GS sequences described herein, and these may be obtained for example by  
35 probing genomic DNA libraries made from such isolates or strains

of bacteria using GS or ORF sequences as probes under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

- A particularly preferred group of pathogenic mycobacteria are isolates of *M.paratuberculosis*. Polynucleotides based on GS regions from such bacteria are particularly preferred. Preferred fragments of such regions include fragments encoding individual open reading frames including the preferred groups and combinations of open reading frames discussed above.
- 10 Alternatively, such polynucleotides may be obtained by site directed mutagenesis of the GS or ORF sequences or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the
- 15 polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides of the invention. Such altered property or function will include the addition of
- 20 amino acid sequences of consensus signal peptides known in the art to effect transport and secretion of the modified polypeptide of the invention. Another altered property will include metagenesis of a catalytic residue or generation of fusion proteins with another polypeptide. Such fusion proteins may be
- 25 with an enzyme, with an antibody or with a cytokine or other ligand for a receptor, to target a polypeptide of the invention to a specific cell type *in vitro* or *in vivo*.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

- 30 Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as <sup>32</sup>P or <sup>35</sup>S, enzyme labels, other protein labels or smaller labels such as biotin or fluorophores. Such labels may be added to polynucleotides or primers of the invention and may be detected
- 35 using by techniques known per se.

- 14 -

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for the presence or absence of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb* applied to samples of body fluids, tissues, or excreta from animals and humans, as well as to food and environmental samples such as river or ground water and domestic water supplies.

Human and animal body fluids include sputum, blood, serum, plasma, saliva, milk, urine, csf, semen, faeces and infected discharges. Tissues include intestine, mouth ulcers, skin, lymph nodes, spleen, lung and liver obtained surgically or by a biopsy technique. Animals particularly include commercial livestock such as cattle, sheep, goats, deer, rabbits but wild animals and animals in zoos may also be tested.

Such tests comprise bringing a human or animal body fluid or tissue extract, or an extract of an environmental or food sample, into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridized to the probe, and then detecting nucleic acid which has hybridized to the probe. Alternatively, the sample nucleic acid may be immobilized on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this any other formats can be found in for example WO89/03891 and WO90/13667.

Polynucleotides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for

- 15 -

which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

- 5 The use of polynucleotides of the invention in the diagnosis of inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polynucleotides may also be used in the prognosis of these diseases. For example, the response of a  
10 human or animal subject in response to antibiotic, vaccination or other therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

- The use of *Mtb* polynucleotides (particularly in the form of  
15 probes and primers) of the invention in the above-described methods form a further aspect of the invention, particularly for the detection, diagnosis or prognosis of *Mtb* infections.

#### B. Polypeptides.

- Polypeptides of the invention include polypeptides in  
20 substantially isolated form encoded by GS. This includes the full length polypeptides encoded by the positive and complementary negative strands of GS. Each of the full length polypeptides will contain one of the amino acid sequences set out in Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and  
25 29. Polypeptides of the invention further include variants of such sequences, including naturally occurring allelic variants and synthetic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, e.g. 80%, 90%, 95% or 98%  
30 amino acid homology (identity) over 30 or more, e.g 40, 50 or 100 amino acids. For example, one group of substantially homologous polypeptides are those which have at least 95% amino acid identity to a polypeptide of any one of Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29 over their entire length.  
35 Even more preferably, this homology is 98%.

Polypeptides of the invention further include the polypeptide sequences of the homologous ORFs of *Mtb*, namely Seq ID Nos. 31, 33, 35, 37 and 39. Unless explicitly specified to the contrary, reference to polypeptides of the invention and their fragments include these *Mtb* polypeptides and fragments, and variants thereof (substantially homologous to said sequences) as defined herein.

Polypeptides of the invention may be obtained by the standard techniques mentioned above. Polypeptides of the invention also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NOS:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39. Such fragments for example of 8, 10, 12, 15 or up to 30 or 40 amino acids may also be obtained synthetically using standard techniques known in the art.

Preferred fragments include those which include an epitope, especially an epitope which is specific to the pathogenicity of the mycobacterial cell from which the polypeptide is derived. Suitable fragments will be at least about 5, e.g. 8, 10, 12, 15 or 20 amino acids in size, or larger. Epitopes may be determined either by techniques such as peptide scanning techniques as described by Geysen et al, *Mol.Immunol.*, 23; 709-715 (1986), as well as other techniques known in the art.

The term "an epitope which is specific to the pathogenicity of the mycobacterial cell" means that the epitope is encoded by a portion of the GS region, or by the corresponding ORF sequences of *Mtb* which can be used to distinguish mycobacteria which are pathogenic by from related non-pathogenic mycobacteria including non-pathogenic species of *M.avium*. This may be determined using routine methodology. A candidate epitope from an ORF may be prepared and used to immunise an animal such as a rat or rabbit in order to generate antibodies. The antibodies may then be used to detect the presence of the epitope in pathogenic mycobacteria and to confirm that non-pathogenic mycobacteria do not contain any proteins which react with the epitope. Epitopes may be linear or conformational.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also  
5 be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

10 Polypeptides of the invention may be modified to confer a desired property or function for example by the addition of Histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell.

Thus, polypeptides of the invention include fusion proteins which  
15 comprise a polypeptide encoding all or part of one or more of an ORF of the invention fused at the N- or C-terminus to a second sequence to provide the desired property or function. Sequences which promote secretion from a cell include, for example the yeast  $\alpha$ -factor signal sequence.

20 A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g.  $^{125}\text{I}$ ,  $^{35}\text{S}$  enzymes, antibodies, polynucleotides and ligands such as biotin. Labelled polypeptides of the  
25 invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample. Polypeptides or labelled polypeptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune  
30 reactivity to said polypeptides in animals and humans using standard protocols.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well, microparticle, dipstick or  
35 biosensor. Such labelled and/or immobilized polypeptides may be

packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Such polypeptides and kits may be used in methods of detection of antibodies or cell mediated immunoreactivity, to the mycobacterial proteins and peptides encoded by the ORFs of the invention and their allelic variants and fragments, using immunoassay. Such host antibodies or cell mediated immune reactivity will occur in humans or animals with an immune system which detects and reacts against polypeptides of the invention. The antibodies may be present in a biological sample from such humans or animals, where the biological sample may be a sample as defined above particularly blood, milk or saliva.

Immunoassay methods are well known in the art and will generally comprise:

- (a) providing a polypeptide of the invention comprising an epitope bindable by an antibody against said mycobacterial polypeptide;
- (b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said polypeptide is formed.

Immunoassay methods for cell mediated immune reactivity in animals and humans are also well known in the art (e.g. as described by Weir et al 1994, J.Immunol Methods 176; 93-101) and will generally comprise

- (a) providing a polypeptide of the invention comprising an epitope bindable by a lymphocyte or macrophage or other cell receptor;
- (b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator to occur; and
- (c) detecting the presence of said cytokine or mediator in the incubate.



Polypeptides of the invention may be made by standard synthetic means well known in the art or recombinantly, as described below.

Polypeptides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise  
5 different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The polypeptides of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits  
10 the polypeptide may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be examined, control reagents, instructions, and the like.

The use of polypeptides of the invention in the diagnosis of  
15 inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polypeptides may also be used in the prognosis of these diseases. For example, the response of a human or animal subject in response to antibiotic or other  
20 therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

The use of *Mtb* polypeptides of the invention in the above-described methods form a further aspect of the invention,  
25 particularly for the detection, diagnosis or prognosis of *Mtb* infections.

Polypeptides of the invention may also be used in assay methods for identifying candidate chemical compounds which will be useful in inhibiting, binding to or disrupting the function of said  
30 polypeptides required for pathogenicity. In general, such assays involve bringing the polypeptide into contact with a candidate inhibitor compound and observing the ability of the compound to disrupt, bind to or interfere with the polypeptide.

There are a number of ways in which the assay may be formatted. For example, those polypeptides which have an enzymatic function may be assayed using labelled substrates for the enzyme, and the amount of, or rate of, conversion of the substrate into a product  
5 measured, e.g by chromatography such as HPLC or by a colourimetric assay. Suitable labels include <sup>35</sup>S, <sup>125</sup>I, biotin or enzymes such as horse radish peroxidase.

For example, the gene product of ORF C is believed to have GDP-mannose dehydratase activity. Thus an assay for inhibitors of the  
10 gene product may utilise for example labelled GDP-mannose, GDP or mannose and the activity of the gene product followed. ORF D encodes a gene related to the synthesis and regulation of capsular polysaccharides, which are often associated with invasiveness and pathogenicity. Labelled polysaccharide  
15 substrates may be used in assays of the ORF D gene product. The gene product of ORF F encodes a protein with putative glucosyl transferase activity and thus labelled amino sugars such as  $\beta$ -1-3-N-acetylglucosamine may be used as substrates in assays.

Candidate chemical compounds which may be used may be natural or  
20 synthetic chemical compounds used in drug screening programmes. Extracts of plants which contain several characterised or uncharacterised components may also be used.

Alternatively, the a polypeptide of the invention may be screened against a panel of peptides, nucleic acids or other chemical  
25 functionalities which are generated by combinatorial chemistry. This will allow the definition of chemical entities which bind to polypeptides of the invention. Typically, the polypeptide of the invention will be brought into contact with a panel of compounds from a combinatorial library, with either the panel  
30 or the polypeptide being immobilized on a solid phase, under conditions suitable for the polypeptide to bind to the panel. The solid phase will then be washed under conditions in which only specific interactions between the polypeptide and individual members of the panel are retained, and those specific members may  
35 be utilized in further assays or used to design further panels of candidate compounds.

- 21 -

For example, a number of assay methods to define peptide interaction with peptides are known. For example, WO86/00991 describes a method for determining mimotopes which comprises making panels of catamer preparations, for example octamers of amino acids, at which one or more of the positions is defined and the remaining positions are randomly made up of other amino acids, determining which catamer binds to a protein of interest and re-screening the protein of interest against a further panel based on the most reactive catamer in which one or more additional designated positions are systematically varied. This may be repeated throughout a number of cycles and used to build up a sequence of a binding candidate compound of interest.

WO89/03430 describes screening methods which permit the preparation of specific mimotopes which mimic the immunological activity of a desired analyte. These mimotopes are identified by reacting a panel of individual peptides wherein said peptides are of systematically varying hydrophobicity, amphipathic characteristics and charge patterns, using an antibody against an antigen of interest. Thus in the present case antibodies against the a polypeptide of the inventoin may be employed and mimotope peptides from such panels may be identified.

#### C. Vectors.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

#### D. Expression Vectors.

- 22 -

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably  
5 linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the  
10 control sequences. Such vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell  
15 transformed or transfected with an expression vector as described above, under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

A further embodiment of the invention provides vectors for the  
20 replication and expression of polynucleotides of the invention, or fragments thereof. The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The  
25 vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector  
30 may also be adapted to be used *in vivo*, for example in a method of naked DNA vaccination or gene therapy. A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention, including the DNA of GS, the  
35 open reading frames thereof and other corresponding ORFs particularly ORFs B, C, E and F from *Mtb*. The cells will be chosen to be compatible with the said vector and may for example be bacterial, yeast, insect or mammalian.

- 23 -

Expression vectors are widely available in the art and can be obtained commercially. Mammalian expression vectors may comprise a mammalian or viral promoter. Mammalian promoters include the metallothionien promoter. Viral promoters include promoters from  
5 adenovirus, the SV40 large T promoter and retroviral LTR promoters. Promoters compatible with insect cells include the polyhedrin promoter. Yeast promoters include the alcohol dehydrogenase promoter. Bacterial promoters include the  $\beta$ -galactosidase promoter.

10 The expression vectors may also comprise enhancers, and in the case of eukaryotic vectors polyadenylation signal sequence downstream of the coding sequence being expressed.

Polypeptides of the invention may be expressed in suitable host cells, for example bacterial, yeast, plant, insect and mammalian  
15 cells, and recovered using standard purification techniques including, for example affinity chromatography, HPLC or other chromatographic separation techniques.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in  
20 order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides or ligands may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of the proteins encoded by the ORFs of the invention in a mycobacterial cell.

25 Polynucleotides of the invention may also be carried by vectors suitable for gene therapy methods. Such gene therapy methods include those designed to provide vaccination against diseases caused by pathogenic mycobacteria or to boost the immune response of a human or animal infected with a pathogenic mycobacteria.

30 For example, Ziegner et al, AIDS, 1995, 9;43-50 describes the use of a replication defective recombinant amphotropic retrovirus to boost the immune response in patients with HIV infection. Such a retrovirus may be modified to carry a polynucleotide encoding a polypeptide or fragment thereof of the invention and the

retrovirus delivered to the cells of a human or animal subject in order to provide an immune response against said polypeptide. The retrovirus may be delivered directly to the patient or may be used to infect cells ex-vivo, e.g. fibroblast cells, which  
5 are then introduced into the patient, optionally after being inactivated. The cells are desirably autologous or HLA-matched cells from the human or animal subject.

Gene therapy methods including methods for boosting an immune response to a particluar pathogen are disclosed generally in for  
10 example WO95/14091, the disclosure of which is incorporated herein by reference. Recombinant viral vectors include retroviral vectors, adenoviral vectors, adeno-associated viral vectors, vaccinia virus vectors, herpes virus vectors and alphavirus vectors. Alpha virus vectors are described in, for example,  
15 WO95/07994, the disclosure of which is incorporated herein by reference.

Where direct administration of the recombinant viral vector is contemplated, either in the form of naked nucleic acid or in the form of packaged particles carrying the nucleic acid this may be  
20 done by any suitable means, for example oral administration or intravenous injection. From  $10^5$  to  $10^8$  c.f.u of virus represents a typical dose, which may be repeated for example weekly over a period of a few months. Administration of autologous or HLA-matched cells infected with the virus may be more convenient in  
25 some cases. This will generally be achieved by administering doses, for example from  $10^5$  to  $10^8$  cells per dose which may be repeated as described above.

The recombinant viral vector may further comprise nucleic acid capable of expressing an accessory molecule of the immune system  
30 designed to increase the immune response. Such a moleclue may be for example and interferon, particularly interferon gamma, an interleukin, for example IL-1 $\alpha$ , IL-1 $\beta$  or IL-2, or an HLA class I or II moleclue. This may be particularly desirable where the vector is intended for use in the treatment of humans or animals  
35 already infected with a mycobacteria and it is desired to boost the immune response.

E. Antibodies.

The invention also provides monoclonal or polyclonal antibodies to polypeptides of the invention or fragments thereof. The invention further provides a process for the production of  
5 monoclonal or polyclonal antibodies to polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using the polypeptides of the invention or peptide fragments thereof, as immunogens. Polyclonal antibodies may also be prepared by conventional means which comprise  
10 inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention or peptide fragment thereof and recovering immune serum.

In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof  
15 haptenised to another polypeptide for use as immunogens in animals or humans.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a polypeptide of the  
20 invention. Such fragments include Fv, F(ab') and F(ab')<sub>2</sub> fragments, as well as single chain antibodies. Furthermore, the antibodies and fragments thereof may be humanised antibodies, e.g. as described in EP-A-239400.

Antibodies may be used in methods of detecting polypeptides of  
25 the invention present in biological samples (where such samples include the human or animal body samples, and environmental samples, mentioned above) by a method which comprises:

- (a) providing an antibody of the invention;
- (b) incubating a biological sample with said antibody  
30 under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said antibody is formed.

- 26 -

Antibodies of the invention may be bound to a solid support for example an immunoassay well, microparticle, dipstick or biosensor and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

- 5 Antibodies of the invention may be used in the detection, diagnosis and prognosis of diseases as described above in relation to polypeptides of the invention.

#### F. Compositions.

- 10 The present invention also provides compositions comprising a polynucleotide or polypeptide of the invention together with a carrier or diluent. Compositions of the invention also include compositions comprising a nucleic acid, particularly and expression vector, of the invention. Compositions further include those carrying a recombinant virus of the invention.
- 15 Such compositions include pharmaceutical compositions in which case the carrier or diluent will be pharmaceutically acceptable.

- Pharmaceutically acceptable carriers or diluents include those used in formulations suitable for inhalation as well as oral, parenteral (e.g. intramuscular or intravenous or transcutaneous)
- 20 administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In
- 25 general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

- For example, formulations suitable for parenteral administration
- 30 include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening



agents, and liposomes or other microparticulate systems which are designed to target the polynucleotide or the polypeptide of the invention to blood components or one or more organs, or to target cells such as M cells of the intestine after oral administration.

5 G. Vaccines.

In another aspect, the invention provides novel vaccines for the prevention and treatment of infections caused by *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria and *Mtb* in animals and humans. The term "vaccine" as used herein means an agent  
10 used to stimulate the immune system of a vertebrate, particularly a warm blooded vertebrate including humans, so as to provide protection against future harm by an organism to which the vaccine is directed or to assist in the eradication of an organism in the treatment of established infection. The immune  
15 system will be stimulated by the production of cellular immunity antibodies, desirably neutralizing antibodies, directed to epitopes found on or in a pathogenic mycobacterium which expresses any one of the ORFs of the invention. The antibody so produced may be any of the immunological classes, such as the  
20 immunoglobulins A, D, E, G or M. Vaccines which stimulate the production of IgA are interest since this is the principle immunoglobulin produced by the secretory system of warm-blooded animals, and the production of such antibodies will help prevent infection or colonization of the intestinal tract. However an  
25 IgM and IgG response will also be desirable for systemic infections such as Crohn's disease or tuberculosis.

Vaccines of the invention include polynucleotides of the invention or fragments thereof in suitable vectors and administered by injection of naked DNA using standard protocols.  
30 Polynucleotides of the invention or fragments thereof in suitable vectors for the expression of the polypeptides of the invention may be given by injection, inhalation or by mouth. Suitable vectors include *M.bovis* BCG, *M.smegmatis* or other mycobacteria, *Corynebacteria*, *Salmonella* or other agents according to  
35 established protocols.

Polypeptides of the invention or fragments thereof in substantially isolated form may be used as vaccines by injection, inhalation, oral administration or by transcutaneous application according to standard protocols. Adjuvants (such as Iscoms or  
5 polylactide-coglycolide encapsulation), cytokines such as IL-12 and other immunomodulators may be used for the selective enhancement of the cell mediated or humoral immunological responses. Vaccination with polynucleotides and/or polypeptides of the invention may be undertaken to increase the susceptibility  
10 of pathogenic mycobacteria to antimicrobial agents *in vivo*.

In instances wherein the polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in  
15 the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridylthio) propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of  
20 cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example,  
25 Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thioether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic  
30 acid, and the like. The carboxyl group can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein  
35 by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

- 29 -

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose®, agarose, cellulose, cellulose beads and the like; 5 polymeric amino acids, such as polyglutamic acid, polylysine, polylactide-coglycolide and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin 10 molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

The immunogenicity of the epitopes may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, 15 that associated with hepatitis B surface antigen. See, e.g., US-A-4,722,840. Constructs wherein the epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the epitope. In addition, all of the vectors prepared include epitopes specific to HBV, 20 having various degrees of immunogenicity, such as, for example, the pre-S peptide.

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an epitope of the invention. In this replacement, regions which are not required 25 to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the epitope of the invention.

Vaccines may be prepared from one or more immunogenic 30 polypeptides of the invention. These polypeptides may be expressed in various host cells (e.g., bacteria, yeast, insect, or mammalian cells), or alternatively may be isolated from viral preparations or made synthetically.

In addition to the above, it is also possible to prepare live 35 vaccines of attenuated microorganisms which express one or more

recombinant polypeptides of the invention. Suitable attenuated microorganisms are known in the art and include, for example, viruses (e.g., vaccinia virus), as well as bacteria.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, or as suitably encapsulated oral preparations and either liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to ingestion or injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween® 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories, oral formulations or as

enemas. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% - 2%. Oral  
5 formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained  
10 release formulations or powders and contain 10% - 95% of active ingredient, preferably 25% - 70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and  
15 which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium,  
20 or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be  
25 prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 $\mu$ g to 250 $\mu$ g, of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, mode of administration and the degree of protection desired.  
30 Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in  
35 which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals

- 32 -

required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the  
5 individual and be dependent upon the judgement of the practitioner.

In a further aspect of the invention, there is provided an attenuated vaccine comprising a normally pathogenic mycobacteria which harbours an attenuating mutation in any one of the genes  
10 encoding a polypeptide of the invention. The gene is selected from the group of ORFs A, B, C, D, E, F, G and H, including the homologous ORFs B, C, E and F in *Mtb*.

The mycobacteria may be used in the form of killed bacteria or as a live attenuated vaccine. There are advantages to a live  
15 attenuated vaccine. The whole live organism is used, rather than dead cells or selected cell components which may exhibit modified or denatured antigens. Protein antigens in the outer membrane will maintain their tertiary and quaternary structures. Therefore the potential to elicit a good protective long term  
20 immunity should be higher.

The term "mutation" and the like refers to a genetic lesion in a gene which renders the gene non-functional. This may be at either the level of transcription or translation. The term thus envisages deletion of the entire gene or substantial portions  
25 thereof, and also point mutations in the coding sequence which result in truncated gene products unable to carry out the normal function of the gene.

A mutation introduced into a bacterium of the invention will generally be a non-reverting attenuating mutation. Non-reverting  
30 means that for practical purposes the probability of the mutated gene being restored to its normal function is small, for example less than 1 in  $10^6$  such as less than 1 in  $10^9$  or even less than 1 in  $10^{12}$ .

An attenuated mycobacteria of the invention may be in isolated form. This is usually desirable when the bacterium is to be used for the purposes of vaccination. The term "isolated" means that the bacterium is in a form in which it can be cultured, processed or otherwise used in a form in which it can be readily identified and in which it is substantially uncontaminated by other bacterial strains, for example non-attenuated parent strains or unrelated bacterial strains. The term "isolated bacterium" thus encompasses cultures of a bacterial mutant of the invention, for example in the form of colonies on a solid medium or in the form of a liquid culture, as well as frozen or dried preparations of the strains.

In a preferred aspect, the attenuated mycobacterium further comprises at least one additional mutation. This may be a mutation in a gene responsible for the production of products essential to bacterial growth which are absent in a human or animal host. For example, mutations to the gene for aspartate semi-aldehyde dehydrogenase (*asd*) have been proposed for the production of attenuated strains of *Salmonella*. The *asd* gene is described further in Gene (1993) 129; 123-128. A lesion in the *asd* gene, encoding the enzyme aspartate  $\beta$ -semialdehyde dehydrogenase would render the organism auxotrophic for the essential nutrient diaminopellic acid (DAP), which can be provided exogenously during bulk culture of the vaccine strain. Since this compound is an essential constituent of the cell wall for gram-negative and some gram-positive organisms and is absent from mammalian or other vertebrate tissues, mutants would undergo lysis after about three rounds of division in such tissues. Analogous mutations may be made to the attenuated mycobacteria of the invention.

In addition or in the alternative, the attenuated mycobacteria may carry a *recA* mutation. The *recA* mutation knocks out homologous recombination - the process which is exploited for the construction of the mutations. Once the *recA* mutation has been incorporated the strain will be unable to repair the constructed deletion mutations. Such a mutation will provide attenuated strains in which the possibility of homologous recombination to

- 34 -

with DNA from wild-type strains has been minimized. RecA genes have been widely studied in the art and their sequences are available. Further modifications may be made for additional safety.

- 5 The invention further provides a process for preparing a vaccine composition comprising an attenuated bacterium according to the invention process comprises (a) inoculating a culture vessel containing a nutrient medium suitable for growth of said bacterium; (b) culturing said bacterium; (c) recovering said  
10 bacteria and (d) mixing said bacteria with a pharmaceutically acceptable diluent or carrier.

- Attenuated bacterial strains according to the invention may be constructed using recombinant DNA methodology which is known per se. In general, bacterial genes may be mutated by a process of  
15 targeted homologous recombination in which a DNA construct containing a mutated form of the gene is introduced into a host bacterium which it is desired to attenuate. The construct will recombine with the wild-type gene carried by the host and thus the mutated gene may be incorporated into the host genome to  
20 provide a bacterium of the present invention which may then be isolated.

- The mutated gene may be obtained by introducing deletions into the gene, e.g by digesting with a restriction enzyme which cuts the coding sequence twice to excise a portion of the gene and  
25 then religating under conditions in which the excised portion is not reintroduced into the cut gene. Alternatively frame shift mutations may be introduced by cutting with a restriction enzyme which leaves overhanging 5' and 3' termini, filling in and/or trimming back the overhangs, and religating. Similar mutations  
30 may be made by site directed mutagenesis. These are only examples of the types of techniques which will readily be at the disposal of those of skill in the art.

- Various assays are available to detect successful recombination. In the case of attenuations which mutate a target gene necessary  
35 for the production of an essential metabolite or catabolite



- 35 -

compound, selection may be carried out by screening for bacteria unable to grow in the absence of such a compound. Bacteria may also be screened with antibodies or nucleic acids of the invention to determine the absence of production of a mutated  
5 gene product of the invention or to confirm that the genetic lesion introduced - e.g. a deletion - has been incorporated into the genome of the attenuated strain.

The concentration of the attenuated strain in the vaccine will be formulated to allow convenient unit dosage forms to be  
10 prepared. Concentrations of from about  $10^4$  to  $10^9$  bacteria per ml will generally be suitable, e.g. from about  $10^5$  to  $10^8$  such as about  $10^6$  per ml. Live attenuated organisms may be administered subcutaneously or intramuscularly at up to  $10^8$  organisms in one or more doses, e.g. from around  $10^5$  to  $10^8$ , e.g. about  $10^6$  or  $10^7$   
15 organisms in a single dose.

The vaccines of the invention may be administered to recipients to treat established disease or in order to protect them against diseases caused by the corresponding wild type mycobacteria, such as inflammatory diseases such as Crohn's disease or sarcoidosis  
20 in humans or Johne's disease in animals. The vaccine may be administered by any suitable route. In general, subcutaneous or intramuscular injection is most convenient, but oral, intranasal and colorectal administration may also be used.

The following Examples illustrates aspects of the invention.

25 **EXAMPLE 1**

Tests for the presence of the GS identifier sequence were performed on 5 $\mu$ l bacterial DNA extracts (25  $\mu$ g/ml to 500  $\mu$ g/ml) using polymerase chain reaction based on the oligonucleotide primers 5'-GATGCCGTGAGGAGGTAAAGCTGC-3' (Seq ID No. 40) and 5'-  
30 GATACGGCTCTTGAATCCTGCACG-3' (Seq ID No. 41) from within the identifier DNA sequences (Seq.ID Nos 1 and 2). PCR was performed for 40 cycles in the presence of 1.5 mM magnesium and an annealing temperature of 58°C. The presence or absence of the correct amplification product indicated the presence or absence

of GS identifier sequence in the corresponding bacterium. GS identifier sequence is shown to be present in all the laboratory and field strains of *Mptb* and *Mavs* tested. This includes *Mptb* isolates 0025 (bovine CVL Weybridge), 0021 (caprine, Moredun), 5 0022 (bovine, Moredun), 0139 (human, Chiodini 1984), 0209, 0208, 0211, 0210, 0212, 0207, 0204, 0206 (bovine, Whipple 1990). All *Mptb* strains were IS900 positive. The *Mavs* strains include 0010 and 0012 (woodpigeon, Thorel) 0018 (armadillo, Portaels) and 0034, 0037, 0038, 0040 (AIDS, Hoffner). All *Mavs* strains were 10 IS902 positive. One pathogenic *M.avium* strain 0033 (AIDS, Hoffner) also contained GS identifier sequence. GS identifier sequence is absent from other mycobacteria including other *M.avium*, *M.malmoense*, *M.szulgai*, *M.gordonae*, *M.chelonei*, *M.fortuitum*, *M.phlei*, as well as *E.coli*, *S.areus*, *Nocardia* sp, 15 *Streptococcus* sp. *Shigella* sp. *Pseudomonas* sp.

#### Example 2:

To obtain the full sequence of GS in *Mavs* and *Mptb* we generated a genomic library of *Mavs* using the restriction endonuclease EcoRI and cloning into the vector pUC18. This achieved a 20 representative library which was screened with <sup>32</sup>P-labelled identifier sequence yielding a positive clone containing a 17kbp insert. We constructed a restriction map of this insert and identified GS as fragments unique to *Mavs* and *Mptb* and not occurring in laboratory strains of *M.avium*. These fragments 25 were sub-cloned into pUC18 and pGEM4Z. We identified GS contained within an 8kb region. The full nucleotide sequence was determined for GS on both DNA strands using primer walking and automated DNA sequencing. DNA sequence for GS in *Mptb* was obtained using overlapping PCR products generated using PwoDNA 30 polymerase, a proofreading thermostable enzyme. The final DNA sequences were derived using the University of Wisconsin GCG gel assembly software package.

#### Example 3:

The DNA sequence of GS in *Mavs* and *Mptb* was found to be more 35 than 99% homologous. The ORFs encoded in GS were identified using GeneRunner and DNASTar computer programmes. Eight ORFs were identified and designated GSA, GSB, GSC, GSD, GSE, GSF, GSG

- 37 -

and GSH. Database comparisons were carried out against the GenEMBL Database release version 48.0 (9/96), using the BLAST and BLIXEM programmes. GSA and GSB encoded proteins of 13.5kDa and 30.7kDa respectively, both of unknown functions. GSC encoded a protein of 38.4kDa with a 65% homology to the amino acid sequence of *rfbD* of *V.cholerae*, a 62% amino acid sequence homology to *gmd* of *E.coli* and a 58% homology to *gca* of *Ps.aeruginosa* which are all GDP-D-mannose dehydratases. Equivalent gene products in *H.influenzae*, *S.dysenteriae*, *Y.enterocolitica*, *N.gonorrhoea*, *K.pneumoniae* and *rfbD* in *Salmonella enterica* are all involved in 'O'-antigen processing known to be linked to pathogenicity. GSD encoded a protein of 37.1kDa which showed 58% homology at the DNA level to *wcaG* from *E.coli*, a gene involved in the synthesis and regulation of capsular polysaccharides, also related to pathogenicity. GSE was found to have a > 30% amino acid homology to *rfbT* of *V.cholerae*, involved in the transport of specific LPS components across the cell membrane. In *V.cholerae* the gene product causes a seroconversion from the Inaba to the Ogawa 'epidemic' strain. GSF encoded a protein of 30.2kDa which was homologous in the range 25-40% at the amino acid level to several glucosyl transferases such as *rfpA* of *K.pneumoniae*, *rfbB* of *K.pneumoniae*, *lgtD* of *H.influenzae*, *lsi* of *N.gonorrhoeae*. In *E.coli* an equivalent gene *galE* adds  $\beta$ -1-3 N-acetylglucosamine to galactose, the latter only found in 'O' and 'M' antigens which are also related to pathogenicity. GSH comprising the ORFs GSH<sub>1</sub> and GSH<sub>2</sub> encodes a protein totalling about 60kDa which is a putative transposase with a 40 - 43% homology at the amino acid level to the equivalent gene product of IS21 in *E.coli*. This family of insertion sequences is broadly distributed amongst gram negative bacteria and is responsible for mobility and transposition of genetic elements. An IS21-like element in *B.fragilis* is split either side of the  $\beta$ -lactamase gene controlling its activation and expression. We programmed an *E.coli* S30 cell-free extract with plasmid DNA containing the ORF GSH under the control of a *lac* promoter in the presence of a <sup>35</sup>S-methionine, and demonstrated the translation of an abundant 60kDa protein. The proteins homologous to GS encoded in other organisms are in general highly antigenic. Thus the proteins encoded by the ORFs

- 38 -

in GS may be used in immunoassays of antibody or cell mediated immuno-reactivity for diagnosing infections caused by mycobacteria, particularly *Mptb*, *Mavs* and *Mtb*. Enhancement of host immune recognition of GS encoded proteins by vaccination using naked specific DNA or recombinant GS proteins, may be used in the prevention and treatment of infections caused by *Mptb*, *Mavs* and *Mtb* in humans and animals. Mutation or deletion of all or some of the ORFs A to H in GS may be used to generate attenuated strains of *Mptb*, *Mavs* or *Mtb* with lower pathogenicity for use as living or killed vaccines in humans and animals. Such vaccines are particularly relevant to Johne's disease in animals, to diseases caused by *Mptb* in humans such as Crohn's disease, and to the management of tuberculosis especially where the disease is caused by multiple drug-resistant organisms.

- 39 -

## SEQUENCE LISTING

Seq. ID No.1

5' - 1 GATCCAACTA AACCCGATGG AACCCCGCGC AAACCTATTGG ACGTCTCCGC GCTACGCAGT  
61 TGGGTTGGCG CCCGCGAATC GCACTGAAAG AGGGCATCGA TGCAACGGTG TCGTGGTACC  
5 121 GCACAAATGC CGATGCCGTG AGGAGGTAAA GCTGCGGGCC GGCCGATGTT ATCCCTCCGG  
181 CCGGACGGGT AGGGCGACCT GCCATCGAGT GGTACGGCAG TCGCCTGGCC GGCGAGGCGC  
241 ATGGCCTATG TGAGTATCCC ATAGCCTGGC TTGGCTCGCC CTAACGCATT ATCAGTTGAC  
301 CGCTTTCCGC CCACGTCGCA GGCTTGCGGC AGCATCCCGT TCAGGTCTCC TCATGGTCCG  
361 GTGTGGCAGC ACCACGCAAG CTCGAACCGA CTCGTTTCCC AATTTGCGAT GCTAATATCG  
10 421 CTCGATGGAT TTTTTCGCGA ACGCCGGCTT GATGGCTCGT AACGTTAGCA CCGAGATGCT  
481 GCGCCACTCC GAACGAAAGC GCCTATTAGT AAACCAAGTC GAAGCATACG GAGTCAACGT  
541 TGTATTGAT GTCGGTGCTA ACTCCGGCCA GTTCGGTAGC GCTTTGCGTC GTGCAGGATT  
601 CAAGAGCCGT ATCGTTTCCT TTGAACCTCT TTCGGGGCCA TTTGCGCAAC TAACGCGCAA  
661 GTCGGCATCG GATC -3'

15 Seq. ID No.2

5' - 1 GATCCGATGC CGACTTGCGC GTTAGTTGCG CAAATGGCCC CGAAAGAGGT TCAAAGGAAA  
61 CGATACGGCT CTTGAATCCT GCACGACGCA AAGCGCTACC GAACTGGCCG GAGTTAGCAC  
121 CGACATCAAT AACAACTTG ACTCCGTATG CTTGACTTG GTTTACTAAT AGGCGCTTTC  
181 GTTCGGAGTG GCGCAGCATC TCGGTGCTAA CGTTACGAGC CATCAAGCCG GCGTTGCGCA  
20 241 AAAAATCCAT CGAGCGATAT TAGCATGCGA AATTGGGAAA CGAGTCGGTT CGAGCTTGCG  
301 TGGTCGTGCC ACACCGGACC ATGAGGAGAC CTGAACGGGA TGCTGCCGCA AGCCTGCGAC  
361 GTGGCGCGAA AGCGGTCAAC TGATAATGCG TAGGGGCGAG CCAAGCCAGG CTATGGGATA  
421 CTCACATAGG CCATGCGCCT CGCCGGCCAG GCGACTGCCG TACCACTCGA TGGCAGGTCG  
481 CCCTACCCGT CCGGCCGGAG GGATAACATC GGCCGGCCCG CAGCTTTACC TCCTCACGGC  
25 541 ATCGGCATTT GTGCGGTACC ACGACACCGT TGCATCGATG CCCTCTTTCA GTGCGATTCT  
601 CGGGCGCCAA CCCAACTGCG TAGCGCGGAG ACGTCCAATA GTTTGCGCGG GGTTCATCTG  
661 GGTTTAGTTG GATC -3'

Seq. ID No.3

1 GAATTCTGGG TTGGAGACGA CGTCGAACTC CTGGTCGGTC TTGCTTCGAA  
51 TGATCGCTGT GATCTGGTCG GCGGTGCCGA CAGGAACCGT CGACTTGTCTG  
101 ACGATCACCT TGTACCGGTC GATGTATGAC CCAATGTCGT CCGCAACCGA  
5 GAAGACGTAC GTCAGGTCCG CCGCCCCGCT TTCACCCATG GCGTCGGGA  
151 CCGCGATGAA AATGACGTCC GCGTGCTCGA TTCGCGTTG CCGGTGGTG  
201 GTGAAGTCAA TCAGCCCGTT CTCACGGTTC CTCGCAATCA ACTCCCAACC  
251 CGGGCTCGAA AATCGGGACA CTGCCTGCGA GGAGCAAATC GATCTTGGCC  
301 TGATCGATAT CGACACAGAC GACATCGTTG CCGCTATCCG CGAGACAGGC  
10 401 GCGCGTGACG AGGCCTACAT AGCCTGATCC GACCACCGAA ATTTTCAAGA  
451 TGACCCCTTC AAGTCCCCGA TCGGTGACG ACCATACTGC CGCAACTCTG  
501 TACCCTCCGT GGGTAATTCG CATGTGCGGT TCGTAAGGAG CAGCCAGCGA  
551 GTCGGGGACG TTCGGTGAGA GAGTCGCAGG ACTACGAGGT TGCCGGTGCG  
601 ATACATCACA GTGTTGCGTC TGTCGGCAAC GATGCAGCAA GAACCCACGG  
15 651 GGCAGCCCTG AACTGCGCGC ATGACCGGTC CTTGTCTGG CACCTTTGAT  
701 CCGCCACCGC TTCCATGCGA ACATGACCGG AATCCATAGC GCGTGGTCAA  
751 GCAGCGGGGA GGTAGACGTC GGTGTCTCT GCTCCAACCG TGTGCGTGAT  
801 AACGATTTCT CTGAACGATC TCGAGGGATT GAAAAGCACC GTGGAGAGCG  
851 TTCGCGCGCA GCGCTATGGG GGGCGAATCG AGCACATCGT CATCGACGGT  
20 901 GGATCGGGCG ACGCCGTCGT GGAGTATCTG TCCGGCGATC CTGGCTTTGC  
951 ATATTGGCAA TCTCAGCCCG ACAACGGGAG ATATGACGCG ATGAATCAGG  
1001 GCATTGCCCA TTCGTCGGGC GACCTGTGTG GGTTTATGCA CTCCACGGAT  
1051 CGTTTCTCCG ATCCAGATGC AGTCGCTTCC GTGGTGGAGG CGCTCTCGGG  
1101 GCATGGACCA GTACGTGATT TGTGGGGTTA CGGGAAAAAC AACCTTGTCTG  
25 1151 GACTCGACGG CAAACCACTT TTCCCTCGGC CGTACGGCTA TATGCCGTTT  
1201 AAGATGCGGA AATTTCTGCT CCGCGCGACG GTTGCATC AGGCGACATT  
1251 CTTGCGCGCG TCGCTGGTAG CCAAGTTGGG CGGTTACGAT CTTGATTTTG  
1301 GACTCGAGGC GGACCAGCTG TTCATCTACC GTGCCGCACT AATACGGCCT  
1351 CCCGTCACGA TCGACCGCGT GGTTCGCGAC TTCGATGTCA CGGGACCTGG  
30 1401 TTCAACCCAG CCCATCCGTG AGCACTATCG GACCCTGCGG CGGCTCTGGG  
1451 ACCTGCATGG CGACTACCCG CTGGGTGGGC GCAGAGTGTC GTGGGCTTAC  
1501 TTGCGTGTGA AGGAGTACTT GATTCGGGCC GACCTGGCCG CATTCAACGC  
1551 GGTAAAGTTC TTGCGAGCGA AGTTCGCCAG AGCTTCGCGG AAGCAAATTT  
1601 CATAGAAACC AACTTCTACT GCCTGACCTG AGCAGCGCCG AGGCGCGCAG  
35 1651 CCGGATCAGT GCGACCTGAA CCGCCAGGTG GAAAGCGCCA CCGATCCCGG  
1701 CACCGAGTGC CTGACGCTTC GGATCCCTTG CACCACAACG AGAGTGAGAG  
1751 CGCCATGATG AGGAAATATC GGCTGGGCGG AGTCAACGCC GGAGTGACAA  
1801 AAGTGAGAAC CCGGTGAAGC GAGCGCTTAT AACAGGGATC ACGGGGCAGG  
1851 ATGGTTCCTA CCTCGCCGAG CTACTACTGA GCAAGGGATA CGAGGTTTAC  
40 1901 GGGCTCGTTC GTCGAGCTTC GACGTTTAA ACCTGCGCGA TCGATCACCT  
1951 CTACGTTGAC CCACACCAAC CCGGCGCGCG CTTGTTCTTG CACTATGCAG  
2001 ACCTCACTGA CCGCACCCCG TTGGTGACCC TGCTCAGCAG TATCGACCCG  
2051 GATGAGGTCT ACAACCTCGC AGCGCAGTCC CATGTGCGCG TCAGCTTTGA  
2101 CGAGCCAGTG CATAACCGAG ACACCACCGG CATGGGATCG ATCCGACTTC  
45 2151 TGGAAGCAGT CCGCCTTTCT CCGGTGGACT GCGGTTCTA TCAGGCTTCC  
2201 TCGTCGGAGA TGTTCGGCGC ATCTCCGCCA CCGCAGAACG AATCGACGCC  
2251 GTTCTATCCC CGTTCGCCAT ACGGCGCGGC CAAGGTCTTC TCGTACTGGA  
2301 CGACTCGCAA CTATCGAGAG GCGTACGGAT TATTCGAGT GAATGGCATC  
2351 TTGTTCAACC ATGAGTCCCC CCGGCGCGGC GAGACTTTTC TGACCCGAAA  
50 2401 GATCACGCGT GCCGTGGCGC GCATCCGAGC TGGCGTCCAA TCGGAGGTCT  
2451 ATATGGGCAA CCTCGATGCG ATCCGCGACT GGGGCTACGC GCCCGAATAT  
2501 GTCGAGGGGA TGTGGAGGAT GTTGAAGCG CCTGAACCTG ATGACTACGT

- 41 -

2551 CCTGGCGACA GGGCGTGGTT ACACCGTACG TGAGTTCGCT CAAGCTGCTT  
2601 TTGACCATGT CGGGCTCGAC TGGCAAAAGC GCGTCAAGTT TGACGACCGC  
2651 TATTTGCGTC CCACCGAGGT CGATTGCTA GTAGGAGATG CCGACAAGGC  
2701 GGCCCAAGTCA CTCGGCTGGA AAGCTTCGGT TCATACTGGT GAAGTTCGCGC  
5 2751 GCATCATGGT GGACGCGGAC ATCGCCGCGT TGGAGTGC GAAGTTCGCGC  
2801 TGGATCGACA CGCCGATGTT GCCTGGTTGG GGCAGAGTAA GTTGACGACT  
2851 ACACCTGGGC CTCTGGACCG CGCAACGCCC GTGTATATCG CCGGTCATCG  
2901 GGGGCTGGTC GGCTCAGCGC TCGTACGTAG ATTTGAGGCC GAGGGGTTCA  
2951 CCAATCTCAT TGTGCGATCA CGCGATGAGA TTGATCTGAC GGACCGAGCC  
10 3001 GCAACGTTTG ATTTTGTGTC TGAGACAAGA CCACAGGTGA TCATCGATGC  
3051 GGCCGCACGG GTCGGCGGCA TCATGGCGAA TAACACCTAT CCCGCGGACT  
3101 TCTTGTCGCA AAACCTCCGA ATCCAGACCA ATTTGCTCGA CGCAGCTGTC  
3151 GCCGTGCGTG TGCCGCGGCT CTTTTCCTC GGTTCGTCAT GCATCTACCC  
3201 GAAGTACGCT CCGCAACCTA TCCACGAGAG TGCTTTATTG ACTGGCCCTT  
15 3251 TGGAGCCAC CAACGACGCG TATGCGATCG CCAAGATCGC CGGTATCCTG  
3301 CAAGTTCAGG CGGTTAGGCG CCAATATGGG CTGGCGTGA TCTCTGCGAT  
3351 GCCGACTAAC CTCTACGGAC CCGGCGACAA CTTCTCCCCG TCCGGGTCGC  
3401 ATCTCTTGCC GGCCTCATC CGTCGATATG AGGAAGCCAA AGCTGGTGGT  
3451 GCAGAAGAGG TGACGAATTG GGGGACCGGT ACTCCGCGGC GCGAACTTCT  
20 3501 GCATGTCGAC GATCTGGCGA GCGCATGCCT GTTCCTTTTG GAACATTTCG  
3551 ATGGTCCGAA CCACGTCAAC GTGGGCACCG GCGTCGATCA CAGCATTAGC  
3601 GAGATCGCAG ACATGGTCCG TACAGCGGTG GGCTACATCG GCGAAACACG  
3651 TTGGGATCCA ACTAAACCCG ATGGAACCCC GCGCAACTA TTGGACGCTT  
3701 CCGCGCTACG CGAGTTGGGT TGGCGCCCGC GAATCGCACT GAAAGACGGC  
25 3751 ATCGATGCAA CGGTGTCGTG GTACCGCACA AATGCCGATG CCGTGAGGAG  
3801 GTAAAGCTGC GGGTCGGCCG ATGTTATCCC TCCGGCCGGA CGGGTGGGGC  
3851 GACCTGCCGT CGAGTGGTAC GGCAGTCGCC TGGCCGCGGA GGCGCGTGGC  
3901 CTATGGGAGT ATCCAATAGC CTGGCTTGGC TCGCCCTAC GCATTATCAG  
3951 TTGACCGCTT TCGCGCCAGC TCGCAGGCTT GCGGCAGCAT CCCGTTCCAGG  
30 4001 TCTCCTCATG GTCCGGTGTG GCACGACCAC GCAAGCTCGA ACCGACTCGT  
4051 TTCCCAATTT CGCATGCTAA TATCGCTCGA TGGATTTTTT GCGCAACGCC  
4101 GGCTTGATGG CTCGTAACGT TAGTACCGAG ATGCTGCGCC ACTTCGAACG  
4151 AAAGCGCTA TTAGTAAACC AATTCAAAGC ATACGGAGTC AACGTTGTTA  
4201 TTGATGTCGG TGCTAACTCC GGCCAGTTCG GTAGCGCTTT GCCTCGTGCA  
35 4251 GGATTCAAGA GCCGTATCGT TTCTTTGAA CCTCTTTCGG GGCCATTTCG  
4301 GCAACTAACG CGCAAGTCGG CATCGGATCC ACTATGGGAG TGTCAACAGT  
4351 ATGCCCTAGG CGACGCCGAT GAGACGATTA CCATCAATGT GGCAGGCAAT  
4401 GCGGGGGCAA GTAGTTCGT GCTGCCGATG CTTAAAAGTC ATCAAGATGC  
4451 CTTTCTCTCC GCGAATTATA TTGGCACC GAACGTTGCA ATACCCGCC  
40 4501 TTGATTCGGT TGCATCAGAA TTTCTGAACC CTACCGATGT TACTTCTCTG  
4551 AAGATCGACG TACAGGGTTT CGAGAAGCAG GTTATCACGG GCAGTAAGTC  
4601 AACGCTTAAC GAAAGCTGCG TCGGCATGCA ACTCGAACTT TCTTTTATTC  
4651 CGTTGTACGA AGGTGACATG CTGATTCATG AAGCGCTTGA ACTTGTCTAT  
4701 TCCCTAGGTT TCAGACTGAC GGGTTTGTG CCCGGCTTTA CGGATCCGCG  
45 4751 CAATGGTTCGA ATGCTTCAAG CTGACGGCAT TTTCTTCCGT GGGGACGATT  
4801 GACATAAATG CTCCGTCCGC ACCCTGCCG TATCCAAACG GGCGATCTGG  
4851 TGAGCCGGCC TCCCGGGCAC CTAATCGACT ATCTAAATTG AGGCGGCCGC  
4901 GACGTGCGGC ACGAACAGGT GGCCGGCTGC TAGCGTTACA CACGTCATGA  
4951 CTGCGCCAGT GTTCTCGATA ATTATCCCTA CCTTCAATGC AGCGGTGACG  
50 5001 CTGCAAGCCT GCCTCGGAAG CATCGTCGGG CAGACCTACC GGAAGTGGG  
5051 AGTGGTCCTT GTCGACGGCG GTTCGACCGA TCGGACCCTC GACATCGCGA  
5101 ACAGTTTCCG CCCGGAATC GGCTCGCGAC TGGTCGTTCA CAGCGGGCCC  
5151 GATGATGGCC CCTACGACGC CATGAACCGC GGCGTCGGCG TGGCCACAGG

- 42 -

5201 CGAATGGGTA CTTTTTTTAG GCGCCGACGA CACCCCTCTAC GAACCAACCA  
5251 CGTTGGCCCA GGTAGCCGCT TTTCTCGGCG ACCATGCGGC AAGCCATCTT  
5301 GTCTATGGCG ATGTTGTGAT GCGTTCGACG AAAAGCCGGC ATGCCGGACC  
5351 TTTGACCTC GACCGCCTCC TATTTGAGAC GAATTTGTGC CACCAATCGA  
5401 TCTTTTACCG CCGTGAGCTT TTCGACGGCA TCGGCCCTTA CAACCTGCGC  
5451 TACCGAGTCT GGGCGGACTG GGAATTCAAT ATTCGCTGCT TCTCCAACCC  
5501 GCGCGTGATT ACCCGCTACA TGGACGTCGT GATTTCGAA TACAACGACA  
5551 TGACCGGCTT CAGCATGAGG CAGGGGACTG ATAAAGAGTT CAGAAAACGG  
5601 CTGCCAATGT ACTTCTGGGT TGCAGGGTGG GAGACTTGCA GCGCATGCT  
5651 GCGGTTTTTG AAAGACAAGG AGAATCGCCG TCTGGCCTTG CGTACGCGGT  
5701 TGATAAGGGT TAAGGCCGTC TCCAAAGAAC GAAGCGAGA ACCGTAGTCG  
5751 CGGATCCACA TTGACTTCT TTAACGCGTT TCGTCTGA TCCACCTTTC  
5801 AAGCCCGTTC CGCGTAACGC GCGCGCAGA GAGTGGTCGC ATATCGCATC  
5851 ACTGTTCTCG TGCCAGTGCT TGGAAAGCGT CGAGCACTCT GGTTCGCGTT  
5901 CTTGACGTTT CCGCCCGCTC CTAGAGGTAG CGTGTCACGT GACTGAAGCC  
5951 AATGAGTGCA ACTCGGCGTC GCGAAAGGTT TCAGTCGCGG TTGAGCAAGA  
6001 CACCGCAAGA CTA CTGGAGT GCGTGACAA GCGCCTCCAG CTCGCGGCTG  
6051 AAAGCGGATG CAAAGGGATT CGAAGCTTGA GCAACATGCG AAGGGGAGAA  
6101 CCGCCTATGA GGCTGGGACA GGTTTTCGAT CCGCGCGCGA ATGCACTGTC  
6151 AATGGCCAAG TAGAAGTCCC CGCTGGTGGC CAGCAGAAGT CCCCACTCCG  
6201 CTGCGGGTGG TTGGCTAATT CTTGGCGGCT CCCTTCTTGT GGTGCGCGTG  
6251 GCGCATCCGG TAGGACTCGC CGGAGGTGAC GACGATGCTG GCGTGGTGCA  
6301 GCAGCCGATC GAGGATGCTG GCGGCGGTGG TGTGCTCGGG CAGGAATCGC  
6351 CCCCATTGTT CGAAGGGCCA ATGCGAGGCG ATGGCCAGGG AGCGGCGCTC  
6401 GTAGCCGGCA GCCACGAGCC GGAACAACAG TTGAGTCCCG GTGTCGTCGA  
6451 GCGGGGCGAA CCGCATCTCG TCCAAGATGA CCAGATCCGC GCGGAGCAGG  
6501 GTGTCGATGA TCTTGCCGAC GGTGTTGTCG GCCAGGCCGC GGTAGAGGAC  
6551 CTCGATCAGG TCGGCGGCGG TGAAGTAGCG GACTTTGAAT CCGGCGTGGA  
6601 CCGCAGCGTG CCCGAGCCG ATGAGCAGGT GACTTTTGCC CGTACCAGGT  
6651 GGGCCAATGA CCGCCAGGTT CTGTTGTGCC CGAATCCATT CCAGGCTCGA  
6701 CAGGTAGTCG AACGTGGCTG CCGTGATCGA CGATCCGGTG ACGTCGAACC  
6751 CGTCGAGGGT CTTGGTGACC GGAAGGCTG CCGCCTTGAG ACGGTTGGCG  
6801 GTGTTGGAGG CATCGCGGCG AGCGATCTCG GCCTCAACCA ACGTCCGCG  
6851 GATCTCCTCC GGTGTCCAGC GTTGCCTCTT GCGGACTTGC AACACCTCGG  
6901 CCGCGTTGCG GCGCACCGTG GCCAGCTTCA ACCGCCGCG CCGCGCTCA  
6951 AGGTGAGCAG CCAGCGGTGC CGCCGAGGAC GGTGCCACCG GCTTGGCAGC  
7001 GGTGGTCATG AGGCCGTCCC GTCGGTGGTG TTGATCTTGT AGGCCCTCAA  
7051 CGAGCGGGTC TCGACGGTGG GCAGATCGAG CACGAGTGCG TCGCCGGCGG  
7101 GCGGGGGTTG TGGGGTGCCG GCGCCGGCGG CCAGGATCGA GCGCACGTGC  
7151 GCAGCGCGGA ACCGGCGAAA CGCAACCGCC CGGCGCAGCG CGTCAATCAA  
7201 AGCCTGTTCC CCGTGGGCGG CGCCAAGGCC GAGCAGAATG TCGAGTTCGG  
7251 ATTTAGTCG GGTGTTGCCG ATCGCAGCAG CACCGACGAG GAACTGCTGC  
7301 GCTTCGGTTC CCAATGCGCA GAATCGTTTC TCTGCTTGGG TTTTCGGGCG  
7351 AGGACCACGC GAGGGTGCGG GTCTGGGTCC GTCGTAGTGT TCATCGAGGA  
7401 TGGACACCTC ACCTGGGCTG ACGAGCTCGT GCTCGGCCAC GATCACACCG  
7451 GTCGAGGTT CCAACAGGAT CAGGGCGCCA TGATCGACCA CCACCGCCAC  
7501 GGTGGCACCG ACGAGCCGCT GAGGCACCGA GTAACGAGCT GAGCCGTAAC  
7551 GGATGCACGA GAGGCCGTCG ACCTTACGGC GCACCGACCC CGAGCCGATC  
7601 GTCGGCCGCA GCGAGGGCAG CTCCTCAAG ACGGTGCGCT CGTCAACCAA  
7651 GCGATCGTTG GGCACGGCGC AGATCTCCGA GTGGACCGTG GCATTGACCT  
7701 CCGCGCACCA TAGTTGCGCC TGGGCGTTGA GGGCACGTAG GTCGACCTGC  
7751 TCACCGGCTA ACGCAGCTTC GGTGAGCAGC GGCACCGCAA GGTGCTCCTG  
7801 AGCGTAGCCA CAGAGGTTCT CCACGATGCC CTTGATTGCG GATCCGCAC



7851 CGTGGCAGAA GTCCGGAACG AAGCCATAGT GGGACGCGAA TCGCACATAA  
7901 TCCGGTGTGTG GAACAACAAC ATTGGCGACG ACACCACCTT TGAGGCAGCC  
7951 CATCCGGTCG GCCAGGATCT TGGCCGGAAC CCCACCGATC GCCTC

Seq. ID No.4

5 1 TTCTACTGCC TGACCTGAGC AGCGCCGAGG CGCGCAGCGC GATCACTGCG ACCTGAATGG  
61 CCAGGTGGAA AGCGCCACCG ATCCCGGCAC CGAGTGCCTG ACGATTCGGA TCCCTTGAC  
121 CACAACGAGA GTGAGACCGC CATGATGACG AAATATCGGC TGGGCGGAGT CAACGCCGGA  
181 GTGACAAAAG TGAGAACCCG GTGAAGCGAG CGCTTATAAC AGGGATCAGC GGGCAGGATG  
241 GTTCCTACCT CGCCGAGCTA CTA CTGAGCA AGGATACGA GGTTACGGG CTCGTTTCGTC  
10 301 GAGCTTCGAC GTTTAACAGC TCGCGGATCG ATCACCTCTA CGTTGACCCA CACCAACCGG  
361 GCGCGCGCTT GTTCTTGAC TATGCAGACC TCACTGACGG CACCCGGTTG GTGACCTGTC  
421 TCAGCAGTAT CGACCCGGAT GAGGTCTACA ACCTCGCAGC GCAGTCCCAT GTGCGCGTCA  
481 GCTTTGACGA GCCAGTGCAT ACCGGAGACA CCACCGGCAT GGGATCGATC CGACTTCTGG  
541 AAGCAGTCCG CCTTTCTCGG GTGGACTGCC GGTCTATCA GGCTTCCTCG TCGGAGATGT  
15 601 TCGGCGCATC TCCGCCACCG CAGAACGAAT CGACGCCGTT CTATCCCGT TCGCCATACG  
661 GCGCGGCCAA GGTCTTCTCG TACTGGACGA CTCGCAACTA TCGAGAGGCG TACGGATTAT  
721 TCGCAGTGAA TGGCATCTTG TTCAACCATG AGTCCCCCGG GCGCGGCGAG ACTTTCGTGA  
781 CCCGAAAGAT CACGCGTGCC GTGGCGCGCA TCCGAGCTGG CGTCCAATCG GAGGTCTATA  
841 TGGGCAACCT CGATGCGATC CGCGACTGGG GCTACGCGCC CGAATATGTC GAGGGGATGT  
20 901 GGAGGATGTT GCAAGCGCCT GAACCTGATG ACTACGTCCT GGCACAGGG CGTGGTTACA  
961 CCGTACGTGA GTTCGCTCAA GCTGCTTTTG ACCACGTCGG GCTCGACTGG CAAAAGCAGC  
1021 TCAAGTTTGA CGACCGCTAT TTGCGCCCCA CCGAGGTCGA TTCGCTAGTA GGAGATGCCG  
1081 ACAGGGCGGC CCAGTCACTC GGCTGGAAAG CTTGCGTTCA TACTGGTGAA CTCGCGCGCA  
1141 TCATGGTGGA CGCGACATC GCGCGTCGG AGTGCGATGG CACACCATGG ATCGACACGC  
25 1201 CGATGTTGCC TGTTGGGGC GGAGTAAGTT GACGACTACA CCTGGGCCTC TGGACCGCGC  
1261 AACGCCCCGT TATATCGCCG GTCATCGGGG GCTGGTCGGC TCAGCGCTCG TACGTAGATT  
1321 TGAGGCCGAG GGGTTCACCA ATCTCATTGT GCGATCACGC GATGAGATTG ATCTGACGGA  
1381 CCGAGCCGCA ACGTTTGATT TTGTGTCTGA GACAAGACCA CAGGTGATCA TCGATGCGGC  
1441 CGCACGGGTC GCGGCATCA TGGCGAATAA CACCTATCCC GCGGACTTCT TGTCCGAAAA  
30 1501 CCTCCGAATC CAGACCAATT TGCTCGACGC AGCTGTCGCC GTGCGTGTGC CGCGGCTCCT  
1561 TTTCTCTCGT TCGTCATGCA TCTACCCGAA GTACGCTCCG CAACCTATCC ACGAGAGTGC  
1621 TTTATTGACT GGCCCTTTGG AGCCCACCAA CGACGCGTAT GCGATCGCCA AGATCGCCGG  
1681 TATCCTGCAA GTTCAGGCGG TTAGGCGCCA ATATGGGCTG GCGTGGATCT CTGCGATGCC  
1741 GACTAACCTC TACGGACCCG GCGACAACCT CTCCCGTCC GGGTCGCATC TCTTGCCGGC  
35 1801 GCTCATCCGT CGATATGAGG AAGCCAAAGC TGGTGGTGCA GAAGAGGTGA CGAATTGGGG  
1861 GACCGGTACT CCGCGGCGCG AACTTCTGCA TGTCGACGAT CTGGCGAGCG CATGCCTGTT  
1921 CCTTTTGAA CATTTCGATG GTCCGAACCA CGTCAACGTG GGCACCGGCG TCGATCACAG  
1981 CATTAGCGAG ATCGCAGACA TGGTCGCTAC GGCGGTGGGC TACATCGGCG AAACACGTTG  
2041 GGATCCAACCT AAACCCGATG GAACCCCGCG CAAACTATTG GACGTCTCCG CGCTACGCGA  
40 2101 GTTGGGTTGG CGCCCGCGAA TCGCACTGAA AGACGGCATC GATGCAACGG TGTGTTGGTA  
2161 CCGCACAAT GCGGATGCCG TGAGGAGGTA AAGCTGCGGG CCGGCCGATG TTATCCCTCC  
2221 GGCCGGACGG GTAGGGCGAC CTGCCATCGA GTGGTACGGC AGTCGCCTGG CCGGCGAGGC  
2281 GCATGGCCTA TGGGAGTATC CCATAGCCTG GCTTGGCTCG CCCCTACGCA TTATCAGTTG  
2341 ACCGCTTTCG CGCCAGCTCG CAGGCTCGCG GCAGCATCCC GTTCAGGTCT CCTCATGGTC  
45 2401 CCGTGTGGCA CGACCAGCA AGCTCGAACC GACTCGTTTC CCAATTTGCG ATGCTAATAT  
2461 CGCTCGATGG ATTTTTTGCG CAACGCCGGC TTGATGGCTC GTAACGTTAG CACCGAGATG  
2521 CTGCGCCACT TCGAACGAAA GCGCCTATTA GTAAACCAAT TCAAAGCATA CGGAGTCAAC  
2581 GTTGTATTG ATGTCGGTGC TAACTCCGGC CAGTTCGGTA GCGCTTTGCG TCGTGCAGGA  
2641 TTCAAGAGCC GTATCGTTTC CTTTGAACCT CTTTCGGGGC CATTTCGCGA ACTAACCGGC  
50 2701 GAGTCGGCAT CGGATCCACT ATGGGAGTGT CACCAGTATG CCCTAGGCGA CGCCGATGAG

- 44 -

2761 ACGATTACCA TCAATGTGGC AGGCAATGCG GGGGCAAGTA GTTCCGTGCT GCCGATGCTT  
 2821 AAAAGTCATC AAGATGCCTT TCCTCCCGCG AATTATATTG GCACCGAAGA CGTTGCAATA  
 2881 CACCGCCTTG ATTCGGTTGC ATCAGAATTT CTGAACCCTA CCGATGTTAC TTTCCCTGAAG  
 2941 ATCGACGTAC AGGGTTTCGA GAAGCAGGTT ATCGCGGGCA GTAAGTCAAC GCTTAACGAA  
 5 3001 AGCTGCGTCG GCATGCAACT CGAACTTTCT TTTATTCCGT TGTACGAAGG TGACATGCTG  
 3061 ATTCATGAAG CGCTTGAAC TGTCTATTCC CTAGGTTTCA GACTGACGGG TTTGTTGCCC  
 3121 GGATTTACGG ATCCGCGCAA TGGTCGAATG CTTCAAGCTG ACGGCATTTT CTTCCGTGGG  
 3181 GACGATTGAC ATAAATGCTT GCGTCGGCAC CCTGCCGGTA TCCAAACGGG CGATCTGGTG  
 3241 AGCCGGCCTC CCGGGCACCT AATCGACTAT CTAAATTGAG GCGGCCGCGA CGTGCGGCAC  
 10 3301 GAACAGGTGG CCGGCTGCTA GCGTTACACA CGTCATGACT GCGCCAGTGT TCTCGATAAT  
 3361 TATCCCTACC TTCAATGCAG CGGTGACGCT GCAAGCCTGC CTCGGAAGCA TCGTCGGGCA  
 3421 GACCTACCGG GAAGTGGAAG TGGTCCTTGT CGACGGCGGT TCGACCGATC GGACCCTCGA  
 3481 CATCGCGAAC AGTTTCCGCC CGGAACCTCG CTGCGACTG GTCGTTTACA GCGGGCCCGA  
 3541 TGATGGCCCC TACGACGCCA TGAACCGCGG CGTCGGCGTA GCCACAGCGG AATGGGTACT  
 15 3601 TTTTTTAGGC GCCGACGACA CCCTCTACGA ACCAACCACG TTGGCCAGG TAGCCGCTTT  
 3661 TCTCGGCGAC CATGCGGCAA GCCATCTTGT CTATGGCGAT GTTGTGATGC GTTCGACGAA  
 3721 AAGCCGGCAT GCCGACCTT TCGACCTCGA CCGCCTCCTA TTTGAGACGA ATTTGTGCCA  
 3781 CCAATCGATC TTTTACCGCC GTGAGCTTTT CGACGGCATC GGCCCTTACA ACCTGCGCTA  
 3841 CCGAGTCTGG GCGGACTGGG ACTTCAATAT TCGCTGCTTC TCCAACCCGG CGCTGATTAC  
 20 3901 CCGCTACATG GACGTCGTGA TTTCCGAATA CAACGACATG ACCGGCTTCA GCATGAGGCA  
 3961 GGGGACTGAT AAAGAGTTCA GAAAACGGCT GCCAATGTAC TTCTGGGTTG CAGGGTGGGA  
 4021 GACTTGCAGG CGCATGCTGG CGTTTTTGAA AGACAAGGAG AATCGCCGTC TGGCCTTGCG  
 4081 TACGCGGTTG ATAAGGGTTA AGGCCGTCTC CAAAGAACGA AGCGCAGAAC CGTAGTCGCG  
 4141 GATCCACATT GGACTTCTTT AACGCGTTTG CGTCCTGATC CACCTTTCAA CCCCCTTCCG  
 25 4201 CGTGACGCGG CGCGCAGAGA GTGGTCGCAT ATCGCGTCAC TGTTCCTGTG CCAGTGCTTG  
 4261 GAAAGCGTCG AGCACTCTGG TTCGCGTTCT TGACGTTTCG CCCCCTCCCT AGAGGTAGCG  
 4321 TGTCACGTGA CTGAAGCCAA TGAGTGCAAC TCGGCGTCGC GAAAGGTTTC AGTCGCGGTT  
 4381 GAGCAAGACA CCGCAAGACT ACTGGAGTGC GTGCACAAGC GCCTCCAGCT CACGG

Seq. ID No.5

30 1 atgatcgctg tgatctggtc ggcggtgccc acaggaaccg tcgacttgct gacgatcacc  
 61 ttgtaccggt cgatgtatga cccaatgtcg tccgcaaccg agaagacgta cgtcaggctc  
 121 gccgccccgc ttccacccat gggcgctcggg acggcgatga aaatgacgct cgcgtgctcg  
 181 attccgcggt gccggtcggg ggtgaagtca atcagcccggt tctcacgggt cctcgcaatc  
 241 aactcccaac ccgggctcga aaatcgggac actgcctgcy aggagcaaat cgatcttggt  
 35 301 ctgatcgata tcgacacaga cgacatcggt gccgctatcc gcgagacagg cgcgcgtgac  
 361 gaggcctaca tagcctga

Seq. ID No.6

40 1 M I A V I W S A V P T G T V D L S T I T L Y R S M Y D P M S  
 31 S A T E K T Y V R S A A P L S P M G V G T A M K M T S A C S  
 61 I P R C R S V V K S I S P F S R F L A L N S Q P G L E N R D  
 91 T A C E E Q I D L G L I D I D T D D I V A A I R E T G A R D  
 121 E A Y I A

- 45 -

## Seq. ID No.7

1 gtgtcatctg ctccaaccgt gtcggtgata acgatttcgc tgaacgatct cgagggattg  
61 aaaagcaccg tggagagcgt tcgcgcgcag cgctatgggg gccgaatcga gcacatcgtc  
121 atcgacggtg gatcgggcga cgccgtcgtg gagtatctgt ccggcgatcc tggctttgca  
5 181 tattggcaat ctcagcccca caacgggaga tatgacgcga tgaatcaggg cattgcccat  
241 tcgtcgggag acctgttctg gtttatgcac tccacggatc gtttctccga tccagatgca  
301 gtcgcttccg tgggtggaggc gctctcgggg catggaccag tacgtgattt gtgggggttac  
361 gggaaaaaca accttgcgtg actcgacggc aaaccacttt tccctcggcc gtacggctat  
421 atgcgcgttta agatgcggaa atttctgctc ggcgcgacgg ttgcgcacga ggcgacattc  
10 481 ttccggcgcgt cgctggtagc caagttgggc ggttacgac ttgatttttg actcgaggcg  
541 gaccagctgt tcatctaccg tgcgcacta atacggcctc ccgtcacgat cgaccgcgtg  
601 gtttgcgact tcgatgtcac gggacctggt tcaaccacgc ccacccgtga gcactatcgg  
661 accctgcggc ggctctggga cctgcacggc gactaccgcg tgggtgggag cagagtgtcg  
721 tgggcttact tgcgtgtgaa ggagtacttg attcggggcg acctggccgc attcaacgcg  
15 781 gtaaagtctt tgcgagcgaa gttcggccga gcttcgcgga agcaaaattc atag

## Seq. ID No.8

1 VSSAPT VSVITISLNDLEGLKSTVESVRAQ  
31 RYGGRIEHIVIDGGSGDAVVEYLSGDPGFA  
61 YWQSQPDNGRYDAMNQGIAHSSGDLWFH  
20 91 STD RFS DPDAVASVVEALS GHG PVRDLWGY  
121 GKN NLVGLDGKPLFPRPYGYMPFKMRKFL  
151 GATVAHQATFFGASLVAKLGGYDLDFGLEA  
181 DQLFIYRAALIRPPVTIDRVVCD F DVTGP  
211 STQPIREHYRTLRLWLHGDYPLGGRRVS  
25 241 WAYLRVKEYLIRADLA AFNAVKFLRAKFAR  
271 ASRKQNS

## Seq. ID No.9

1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta  
61 ctactgagca agggatacga ggttcacggg ctcgttcgtc gagcttcgac gtttaacacg  
121 tcgcggatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttgcac  
181 tatgcagacc tcaactgacg caccgggttg gtgacctgc tcagcagtat cgaccgggat  
241 gaggtctaca acctcgacg gcagtcctcat gtgcgcgtca gctttgacga gccagtgcat  
301 accggagaca ccaccggcat gggatcgatc cgacttctgg aagcagtcgg cctttctcgg  
361 gtggactgcc ggttctatca ggcttctcgc tcggagatgt tcggcgcatc tccgccaccg  
35 421 cagaacgaat cgacgcgctt ctatccccgt tcgccatacg gcgcggccaa ggtcttctcg  
481 tactggacga ctcgcaacta tcgagaggcg tacggattat tcgcagtgaa tggcatcttg  
541 ttcaaccatg agtccccccg gcgcggcgag actttcgtga cccgaaagat cagcgtgccc  
601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatgcgac  
661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgcc  
40 721 gaacctgatg actacgtcct ggcgacaggc cgtgggtaca ccgtacgtga gttcgtcaca  
781 gctgcttttg accatgtcgg gctcgactgg caaaagcgcg tcaagtttga cgaccgctat  
841 ttgcgtccca ccgaggtcga ttcgctagta ggagatgccc acaaggcgcc ccagtcactc  
901 ggctggaaag cttcgggtta tactgggtgaa ctgcgcgcga tcatgggtga cgcggacatc  
961 gccgcgttgg agtgcgatgg cacaccatgg atcgacacgc cgatgttgcc tgggtggggc  
45 1021 agagtaagtt ga

- 46 -

## Seq. ID No.10

1 VKRALITGITGQDGSYLAELLLSKGYEV . G  
31 LVRRASTFNTSRIDHLYVDPHQPGARLFLH  
61 YADLTDGTRLVTLSSIDPDEVYNLAAQSH  
5 91 VRVSFDEPVHTGDTTGMGSI RLL EAVRLSR  
121 VDCRFYQASSSEMFGASPPPQNESTPFYPR  
151 SPYGAAKVF SYWTT RNYREAYGLFAVNGIL  
181 FNHESPRRGETFVTRKITRAVARIRAGVQS  
211 EYMG NLD AIRDWGYAPEYVEGMWRMLQAP  
10 241 EPDDYVLATGRGYTVREF AQA AF DHVGLDW  
271 QKRVKFDDRYLRPTEVDSL VG DADKAAQSL  
301 GWKASVHTGELARIMVDADIAALECDGTPW  
331 IDTPMLPGWGRVS

## Seq. ID No.11

15 1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta  
61 ctactgagca agggatacga ggttcacggg ctcgttcgtc gagcttcgac gtttaacacg  
121 tcgcggatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttgac  
181 tatgcagacc tcaactgacg caccgggttg gtgaccctgc tcagcagtat cgaccggat  
241 gaggtctaca acctcgagc gcagtcctcat gtgcgcgtca gctttgacga gccagtgc  
20 301 accggagaca ccaccggcat gggatcgatc cgacttctgg aagcagtcgg cctttctcgg  
361 gtggactgcc ggttctatca ggttcctcgc tcggagatgt tcggcgcatc tccgccaccg  
421 cagaacgaat cgacgcccgt ctatccccgt tcgccatadg gcgcggccaa ggtcttctcg  
481 tactggacga ctcgcaacta tcgagaggcg tacggattat tcgcagtgaa tggcatcttg  
541 ttcaaccatg agtccccccg gcgcggcgag actttcgtga cccgaaagat cagcgtgcc  
25 601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatcgatc  
661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgcc  
721 gaacctgatg actacgtcct ggcgacaggg cgtgggttaca ccgtacgtga gttcgtctaa  
781 gctgcttttg accacgtcgg gctcgactgg caaaagcacg tcaagtttga cgaccgctat  
841 ttgcgcccc cggaggtcga ttcgctagta ggagatgcc acaggggcgc ccagtcactc  
30 901 ggctggaaag cttcgggttca tactggtgaa ctcgcgcgca tcatggtgga cgcggacatc  
961 gccgcgtcgg agtgcgatgg cacaccatgg atcgacacgc cgatgttgcc tgggtggggc  
1021 ggagtaagtt ga

## Seq. ID No.12

1 VKRALITGITGQDGSYLAELLLSKGYEVHG  
35 31 LVRRASTFNTSRIDHLYVDPHQPGARLFLH  
61 YADLTDGTRLVTLSSIDPDEVYNLAAQSH  
91 VRVSFDEPVHTGDTTGMGSI RLL EAVRLSR  
121 VDCRFYQASSSEMFGASPPPQNESTPFYPR  
151 SPYGAAKVF SYWTT RNYREAYGLFAVNGIL  
40 181 FNHESPRRGETFVTRKITRAVARIRAGVQS  
211 EYMG NLD AIRDWGYAPEYVEGMWRMLQAP  
241 EPDDYVLATGRGYTVREF AQA AF DHVGLDW  
271 QKHVKFDDRYLRPTEVDSL VG DADRAA QSL  
301 GWKASVHTGELARIMVDADIAASECDGTPW  
45 331 IDTPMLPGWGGVS

- 47 -

## Seq. ID No.13

1 gtgcgatggc acaccatgga tcgacacgcc gatgttgccct gggtggggca gagtaagttg  
61 acgactacac ctgggcctct ggaccgcgca acgcccggtg atatcgccgg tcatcggggg  
121 ctggtcggct cagcgctcgt acgtagattt gagggcgagg gggtcaccaa tctcattgtg  
5 181 cgatcacgcg atgagattga tctgacggac cgagccgcaa cgtttgattt tgtgtctgag  
241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggcgaataac  
301 acctatcccc cggacttctt gtccgaaaac ctccgaatcc agaccaattt gctcgacgca  
361 gctgtcgcg tgcgtgtgcc gcggctcctt ttccctcggt cgtcatgcat ctaccggaag  
421 tacgtccgc aacctatcca cgagagtget ttattgactg gccctttgga gccaccaaac  
10 481 gacgcgratg cgatcgccaa gatcgccggt atcctgcaag ttcaggcggt taggcgccaa  
541 tatgggctgg cgtggatctc tgcgatgcg actaacctct acggaccggy cgacaacttc  
601 tccccgtccg ggtcgcctct cttgcggcgy ctcacccgct gatatgagga agccaaagct  
661 ggtggtgcag aagaggtgac gaattggggg accggtactc cgcggcgcca acttctgcat  
721 gtcgacgac tggcgagcgc atgcctgttc cttttggaac atttcgatgg tccgaaccac  
15 781 gtcaacgtgg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctaca  
841 gcgggtgggt acatcggcga aacacgttgg gatccaacta aaccgatgg aacccgcgc  
901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccgcgaat cgcactgaaa  
961 gacggcatcg atgcaacggt gtcgtggtac cgcacaaatg ccgatgccgt gaggaggtaa

## Seq. ID No.14

1 V R W H T M D R H A D V A W L G Q S K L T T T P G P L D R A  
31 T P V Y I A G H R G L V G S A L V R R F E A E G F T N L I V  
61 R S R D E I D L T D R A A T F D F V S E T R P Q V I I D A A  
91 A R V G G I M A N N T Y P A D F L S E N L R I Q T N L L D A  
121 A V A V R V P R L L F L G S S C I Y P K Y A P Q P I H E S A  
15 151 L L T G P L E P T N D A Y A I A K I A G I L Q V Q A V R R Q  
181 Y G L A W I S A M P T N L Y G P G D N F S P S G S H L L P A  
211 L I R R Y E E A K A G G A E E V T N W G T G T P R R E L L H  
241 V D D L A S A C L F L L E H F D G P N H V N V G T G V D H S  
271 I S E I A D M V A T A V G Y I G E T R W D P T K P D G T P R  
30 301 K L L D V S A L R E L G W R P R I A L K D G I D A T V S W Y  
331 R T N A D A V R R

- 48 -

Seq. ID No.15

1 gtgcgatggc acaccatgga tcgacacgcc gatgttgccct gggtggggcg gagtaagttg  
 61 acgactacac ctgggcctct ggaccgcgca acgcccggtg atacgcgcgg tcacgsgggg  
 121 ctggctcggt cagcgctcgt acgtagattt gaggcgagg gggtcaccaa tctcattgtg  
 181 cgatcacgcg atgagattga tctgacggac cgagccgcaa cgtttgattt tgtgtctgag  
 241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggcaataaac  
 301 acctatcccg cggacttctt gtccgaaaac ctccgaatcc agaccaatct gctcgacgca  
 361 gctgtcgccg tgcgtgtgcc gcggctcctt ttcctcggtt cgatcatgcat ctaccgaag  
 421 tacgctccgc aacctatcca cgagagtgtt ttattgactg gccctttgga gccaccaaac  
 481 gacgcgtatg cgatcgccaa gatcgccggt atcctgcaag ttcaggcggt taggcgccaa  
 541 tatgggctgg cgtggatctc tgcgatgccg actaacctct acggaccgag cgacaacttc  
 601 tccccgtccg ggtcgatctt cttgccggcg ctcatccgtc gatatgagga agccaaagct  
 661 ggtggtgcag aagaggtgac gaattggggg accggtactc cgcggcgcca acttctgcat  
 721 gtcgacgacg tggcgagcgc atgctgttct cttttggaac atttcgatgg tccgaaccac  
 781 gtcacacgtg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctacg  
 841 gcggtgggct acatcgccga aacacgttgg gatccaaacta aaccgatgg aacccgcgc  
 901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccgcgaaat cgcactgaaa  
 961 gacggcatcg atgcaacggt gtcgtggtac cgcacaaatg ccgatgccgt gaggaggtaa

Seq. ID No.16

1 V R W H T M D R H A D V A W L G R S K L T T T P G P L D R A  
 31 T P V Y I A G H R G L V G S A L V R R F E A E G F T N L I V  
 61 R S R D E I D L T D R A A T F D F V S E T R P Q V I I D A A  
 91 A R V G G I M A N N T Y P A D F L S E N L R I Q T N L L D A  
 121 A V A V R V P R L L F L G S S C I Y P K Y A P Q P I H E S A  
 151 L L T G P L E P T N D A Y A I A K I A G I L Q V Q A V R R Q  
 181 Y G L A W I S A M P T N L Y G P G D N F S P S G S H L L P A  
 211 L I R R Y E E A K A G G A E E V T N W G T G T P R R E L L H  
 241 V D D L A S A C L F L L E H F D G P N H V N V G T G V D H S  
 271 I S E I A D M V A T A V G Y I G E T R W D P T K P D G T P R  
 301 K L L D V S A L R E L G W R P R I A L K D G I D A T V S W Y  
 331 R T N A D A V R R

Seq. ID No.17

1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagtaccga gatgctgcgc  
 61 cacttcgaac gaaagcgccct attagtaaac caattcaaag catacggagt caacgttggt  
 121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggattcaag  
 181 agccgtatcg tttcctttga acctctttcg gggccatttg cgcaactaac gcgcaagtcg  
 241 gcatcgatc cactatggga gtgtcaccag tatgccttag gcgacgccga tgagacgatt  
 301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgctgccgat gcttaaaagt  
 361 catcaagatg cctttctctc cgcgaattat attggcaccg aagacgttgc aatacaccgc  
 421 cttgattcgg ttgcatcaga atttctgaac cctaccgatg ttactttcct gaagatcgac  
 481 gtacaggggt tcgagaagca ggttatcacg ggcagtaagt caacgcttaa cgaaagctgc  
 541 gtcggcatgc aactcgaact ttcttttatt ccgttgtacg aaggtgacat gctgattcat  
 601 gaagcgcttg aacttgtcta ttccttaggt ttcagactga cgggtttggt gcccggttt  
 661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca ttttcttccg tggggacgat  
 721 tga

## Seq. ID No.18

1 M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N  
31 Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K  
61 S R I V S F E P L S G P F A Q L T R K S A S D P L W E C H Q  
5 91 Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S  
121 H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N  
151 P T D V T F L K I D V Q G F E K Q V I T G S K S T L N E S C  
181 V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G  
211 F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D

## 10 Seq. ID No.19

1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagcaccga gatgctgcgc  
61 cacttcgaac gaaagcgect attagtaaac caattcaaag catacggagt caacgttggt  
121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggattcaag  
15 181 agcgcgtatcg tttcctttga acctctttcg gggccatttg cgcaactaac gcgcgagtcg  
241 gcatcggtac cactatggga gtgtcaccag tatgccctag gcgacgccga tgagacgatt  
301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgcgtccgat gcttaaaagt  
361 catcaagatg cctttctctc cgcgaattat attggcaccg aagacgttgc aatacaccgc  
421 cttgattcgg ttgcatcaga atttctgaac cctaccgatg ttactttcct gaagatcgac  
481 gtacaggggt tgcagaagca gggtatcgcg ggcagtaagt caacgcttaa cgaaagctgc  
20 541 gtcggcatgc aactcgaact ttcttttatt ccgttgtacg aaggtgacat gctgattcat  
601 gaagcgcttg aacttgtcta ttccctaggt ttcagactga cgggtttggt gcccggtatt  
661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca ttttcttccg tggggacgat  
721 tga

## Seq. ID No.20

25 1- M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N  
31 Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K  
61 S R I V S F E P L S G P F A Q L T R E S A S D P L W E C H Q  
91 Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S  
121 H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N  
30 151 P T D V T F L K I D V Q G F E K Q V I A G S K S T L N E S C  
181 V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G  
211 F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D

- 50 -

Seq. ID No.21

1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggg gacgctgcaa  
 61 gcttgccctcg gaagcatcgt cgggcagacc taccgggaag tggaagtggg ccttgtcgac  
 121 ggcgggttcga ccgatcggac cctcgacatc gcgaacagtt tccgcccggg actcgggtcg  
 181 cgactggtcg ttcacagcgg gcccgatgat ggcccctacg acgccatgaa ccgcggcgctc  
 241 ggcgtggcca caggcgaatg ggtacttttt ttaggcgccg acgacaccct ctacgaacca  
 301 accacgttgg ccaggttagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat  
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc  
 421 ctctattttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac  
 481 ggcacgcggc cttacaacct gcgctaccga gtctgggcgg actgggactt caatattcgc  
 541 tgcttctcca acccggcgct gattaccgcg tacatggacg tcgtgatttc cgaatacaac  
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca  
 661 atgtacttct ggggttcagg gtgggagact tgcaggcgca tgctggcggt tttgaaagac  
 721 aaggagaatc gccgtctggc cttgcgtacg cgggtgataa gggttaaggc cgtctccaaa  
 781 gaacgaagcg cagaaccgta g

Seq. ID No.22

1 M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T  
 31 Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S  
 61 R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F  
 91 L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y  
 121 G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q  
 151 S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R  
 181 C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G  
 211 T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D  
 241 K E N R R L A L R T R L I R V K A V S K E R S A E P

Seq. ID No.23

1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggg gacgctgcaa  
 61 gcttgccctcg gaagcatcgt cgggcagacc taccgggaag tggaagtggg ccttgtcgac  
 121 ggcgggttcga ccgatcggac cctcgacatc gcgaacagtt tccgcccggg actcgggtcg  
 181 cgactggtcg ttcacagcgg gcccgatgat ggcccctacg acgccatgaa ccgcggcgctc  
 241 ggcgtagcca caggcgaatg ggtacttttt ttaggcgccg acgacaccct ctacgaacca  
 301 accacgttgg ccaggttagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat  
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc  
 421 ctctattttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac  
 481 ggcacgcggc cttacaacct gcgctaccga gtctgggcgg actgggactt caatattcgc  
 541 tgcttctcca acccggcgct gattaccgcg tacatggacg tcgtgatttc cgaatacaac  
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca  
 661 atgtacttct ggggttcagg gtgggagact tgcaggcgca tgctggcggt tttgaaagac  
 721 aaggagaatc gccgtctggc cttgcgtacg cgggtgataa gggttaaggc cgtctccaaa  
 781 gaacgaagcg cagaaccgta g



## Seq. ID No.24

1 M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T  
31 Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S  
61 R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F  
5 91 L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y  
121 G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q  
151 S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R  
181 C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G  
211 T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D  
10 241 K E N R R L A L R T R L I R V K A V S K E R S A E P

## Seq. ID No.25

1 gtggccagca gaagtcccca ctccgctgcg ggtggttggc taattcttgg cggtccctt  
61 cttgtggtcg gcgtggcgca tccggtagga ctccgaggag gtgacgacga tgctggcggtg  
121 gtgcagcagc cgatcgagga tgctggcggc ggtggtgtgc tcgggcagga atcgcccca  
15 181 ttgttcgaag ggccaatgag aggcgatggc caggagcgcg cgctcgtagc cggcagccac  
241 gagccggaac aacagttgag tcccggtgtc gtcgagcggg gcgaagcga tctcgtccaa  
301 gatgaccaga tccgcgcgga gcagggtgtc gatgatcttg ccgacgggtg tgctggccag  
361 gccgcggtag aggacctcga tcaggtcggc gccggtgaag tagcggactt tgaatccggc  
421 gtggacggca gcgtgcccgc agccgatgag cagggtgactt ttgcccgtag cagggtgggc  
20 481 aatgaccgcc aggttctgtt gtgcccgaat ccattccagg ctgcacaggt agtcgaacgt  
541 ggctgcggtg atcgacgacg cggtgacgtc gaaccgctcg agggctcttg tgaccgggaa  
601 ggctgcggcc ttgagacggt tggcggtgtt ggaggcatcg cgggcagcga tctcggcctc  
661 aaccaacgtc cgcaggatct cctccggtgt ccagcgttgc gtcttggcga cttgcaaac  
721 ctccggcgcg ttgcccgcga ccgtggccag cttcaaccgc cgcagcgccg cgtcaaggtc  
25 781 agcagccagc ggtgccgccc aggcaggtgc caccggcttg gcagcgggtg tcatgaggcc  
841 gtcccgtcgg tgggtgtgat cttgtag

## Seq. ID No.26

1 V A S R S P H S A A G G W L I L G G S L L V V G V A H P V G  
31 L A G G D D D A G V V Q Q P I E D A G G G V L G Q E S P P  
61 L F E G P M R G D G Q G A A L V A G S H E P E Q Q L S P G V  
30 91 V E R G E A D L V Q D D Q I R A E Q G V D D L A D G V V G Q  
121 A A V E D L D Q V G G G E V A D F E S G V D G S V P A A D E  
151 Q V T F A R T R W A N D R Q V L L C P N P F Q A R Q V V E R  
181 G C G D R R S G D V E P V E G L G D R E G C G L E T V G G V  
35 211 G G I A G S D L G L N Q R P Q D L L R C P A L R L G D L Q H  
241 L G G V A A H R G Q L Q P P Q R R V K V S S Q R C R R G R C  
271 H R L G S G G H E A V P S V V L I L

- 52 -

Seq. ID No.27

1 atggggctgcc tcaaagggtgg tgctcgtcgcc aatgttggtg ttccaacacc ggattatgtg  
61 cgattcgcgt cccactatgg cttcgttccg gacttctgcc acggtgcgga tccgcaatcg  
121 aagggcatcg tggagaacct ctgtggctac gctcaggacg accttgccgt gccgctgctg  
5 181 accgaagctg cgttagccgg tgagcaggtc gacctacgtg ccttcaacgc ccaggcgcaa  
241 ctatgggtcg ccgaggtcaa tgccacggtc cactcggaga tctgcgccgt gcccaacgat  
301 cgcttggttg acgagcgcac cgtcttgagg gagctgccct cgctgcggcc gacgatcggc  
361 tcgggggtcgg tgcgccgtaa ggtcgacggc ctctcgtgca tccgttacgg ctacagctcgt  
421 tactcgggtg ctcagcggct cgctcgtgcc accgtggcgg tgggtggtcga tcatggcgcc  
10 481 ctgatcctgt tggaaacctg gaccggtgtg atcgtggcgg agcacgagct cgtcagccca  
541 ggtgaggtgt ccatcctcga tgaacactac gacggaccca gaccgcacc ctcgctggtt  
601 cctcgcccgga aaacccaagc agagaaacga ttctgcgcat tgggaaccga agcgcagcag  
661 ttectcgtcg gtgctgctgc gatcggaac acccgactga aatccgaact cgacattctg  
721 ctgggcttg gcgcccga cgccgaacag gctttgattg acgcgctgcg ccggggcggtt  
15 781 gcgtttcggc ggttccggcg tgccgacgtg cgctcgatcc tggccgcccg cgccggcacc  
841 ccacaacccc gccccgcgg cgacgcactc gtgctcgatc tgcccaccgt cgagaccgcg  
901 tcgttgaggg cctacaagat caacaccacc gacgggacgg cctcatgacc accgctgcca  
961 agccgggtgg accgtcctcg gcggcacggc tggctgctga ccttgacgcg gcgctgcggc  
1021 ggttgaagct ggccacggtg cgccgcaacg ccgcccagggt gttgcaagtc gccaaagcgc  
20 1081 aacgctggac accggaggag atcctgcgga cgttggttga ggccgagatc gctgcccgcg  
1141 atgcctccaa caccgccaac cgtctcaagg ccgcagcctt cccggtcacc aagaccctcg  
1201 acgggttcga cgtcaccgga tcgtcgatca ccgcagccac gttcgactac ctgtcgagcc  
1261 tggaaatgat tcgggcacaa cagaacctgg cggtcattgg cccacctggt acggggcaaaa  
1321 gtcacctgct catcggtcgc gggcacgctg ccgtccacgc cggattcaaa gtcgctact  
25 1381 tcaccgcccg cgacctgac gaggtcctct acccgggcct ggccgacaac accgtcggga  
1441 agatcatcga caccctgctc cgcggggatc tggctcatct ggacgagatc ggcttcgccc  
1501 cgctcgacga caccgggact caactgttgt tccggctcgt ggctgcccgc tacgagcgcc  
1561 gctccctggc catcgccctg cattggccct tcgaacaatg ggggagatc ctgcccagagc  
1621 acaccaccgc cgccagcatc ctcgatcgcc tgcgtcacca cgccagcatc gtcgtcacct  
30 1681 ccggcgagtc ctaccggatg cgccacgccc accacaagaa gggagccgcc aagaattag

Seq. ID No.28

1 M G C L K G G V V A N V V V P T P D Y V R F A S H Y G F V P  
31 D F C H G A D P Q S K G I V E N L C G Y A Q D D L A V P L L  
61 T E A A L A G E Q V D L R A L N A Q A Q L W C A E V N A T V  
91 H S E I C A V P N D R L V D E R T V L R E L P S L R P T I G  
121 S G S V R R K V D G L S C I R Y G S A R Y S V P Q R L V G A  
151 T V A V V V D H G A L I L L E P A T G V I V A E H E L V S P  
181 G E V S I L D E H Y D G P R P A P S R G P R P K T Q A E K R  
211 F C A L G T E A Q Q F L V G A A A I G N T R L K S E L D I L  
241 L G L G A A H G E Q A L I D A L R R A V A F R R F R A A D V  
271 R S I L A A G A G T P Q P R P A G D A L V L D L P T V E T R  
301 S L E A Y K I N T T D G T A S

- 53 -

Seq. ID No.29

1 MTTAAKPVAPSSAAPLAADLDAALRRRLKLA  
31 TVRRNAAEVLQVAKTQRWTP EEILRTLVEA  
61 EIAARDASNTANRLKAAAFPVTKTLDGFDV  
5 91 TGSSITAATFDYLS SLEWIRAQQNLAVIGP  
121 PGTGKSHLLIGCGHAAVHAGFKVRYFTAAD  
151 LIEVLYRGLADNTVGKI IDTLLRADLVILD  
181 EIGFAPLDDDTGTQLLFRLVAAGYERRSLAI  
211 ASHWPF EQWGRFLPEHTTAASILDRLLHHA  
10 241 SIVVTSGESYRMRHADHKKGA AKN

Seq. ID No.30

1 gtgacgtctg ctccgaccgt ctccggtgata acgatctcgt tcaacgacct cgacgggttg  
61 cagcgcacgg tga aaagtgt gcgggcgcaa cgctaccggg gacgcacga gcacatcgta  
121 atcgacggtg gcagcgcgca cgacgtggtg gcatacctgt cggggtgtga accaggcttc  
15 181 gcgtattggc agtccgagcc cgacggcggg cggtacgacg cgatgaacca gggcatcgcg  
241 cagcatcggt gtgatctgtt gtggttcttg cactccgccc atcgtttttc cgggcccgc  
301 gtggtagccc aggcctgtga ggcgtatcc ggcaaggagc cgggtgtccga attgtggggc  
361 ttccgggatgg atcgctcgt cgggctcgat cgggtgcgg gcccgatacc ttccagcctg  
421 cgcaaattcc tggccggcaa gcagggtgtt ccgcacaaag catcgttctt cggatcatcg  
20 481 ctggtggcca agatcggtgg ctacgacctt gatttcggga tcccgccga ccaggaattc  
541 atattgcggg ccgcgctggt atcgagccg gtcacgattc ggtgtgtgct gtgcgagttc  
601 gacaccacgg gcgtcggtc gcaccgggaa ccaagcgagg tcttcggtga tctgcgcgc  
661 atgggcgacc ttcacgccc ctaccgcttc gggggaaggc gaatatcaca tgcctacct  
721 cgcggccggg agttctacgc ctacaacagt cgattctggg aaaacgtctt cagcgaatg  
25 781 tcgaaatag

Seq. ID No.31

1 MTSAPT VSVITISFNDLDGLQRTVKS VRAQ  
31 RYRGRIEHIVIDGGSGDDV VAYLSGCEPGF  
61 AYWQSEPDGGRYDAMNQGIAHASGDLLWFL  
30 91 HSADRFSGPDVVAQAVEALSGKGPVSELWG  
121 FGMDRLVGLDRVRGP I PFSLRKFLAGKQVV  
151 PHQASFFGSSLVAKIGGYDLDFGIAADQEF  
181 ILRAALVCEPVTIRCVLCEFDTTGVGSHRE  
211 PSAVFGDLRRMGDLHRRYPFGGRRI SHAYL  
35 241 RGREFYAYNSRFWENVFTRMSK

## Seq. ID No.32

1 gtgaagcgag cgctcatcac cggaatcacc ggccaggacg gctcgtatct cgccgaactg  
 61 ctgctggcca aggggtatga ggttcacggg ctcatccggc gcgcttcgac gttcaacacc  
 121 tcgcggtacg atcacctcta cgtcgaccgg caccaaccgg gcgogcggtt gtttctgcac  
 5 181 tatggtgacc tgatcgacgg aaccgggttg gtgacctgc tgagcaccat cgaaccggac  
 241 gaggtgtaca acctggcggc gcagtcacac gtgcgggtga gcttcgacga acccggtgcac  
 301 accggtgaca ccaccggcat gggatccatg cgaactgctg aagccgttcg gctctctcgg  
 361 gtgcactgcc gcttctatca ggcttcctcg tcggagatgt tcggcgccct gccgccaccg  
 421 cagaacgagc tgacggcgtt ctaccggcg gcaccgtatg gcgccgcaa ggtctattcg  
 10 481 tactgggcga cccgcaatta tcggaagcg tacggattgt tcgccgttaa cggcatcttg  
 541 ttcaatcacg aatcacccgg gcgcggtgag acgttcgtga cccgaaagat caccagggcc  
 601 gtggcacgca tcaaggccgg tatccagtcc gaggtctata tgggcaatct ggtatcggtc  
 661 cgcgactggg ggtacggcgc cgaatacgtc gaaggcatgt ggcggatgct gcagaccgac  
 721 gagcccgacg acttcgtttt ggcgaccggg cgcggtttca ccggtcgtga gttcgcggcg  
 15 781 gccgcgttcg agcatgccgg tttggactgg cagcagtagc tgaaattcga ccaacgctat  
 841 ctgcggccca ccgaggtgga ttcgctgac gccgacgga ccaaggctgc cgaattgctg  
 901 ggctggaggg cttcggtgca cactgacgag ttggctcgga tcatggtcga cgcggacatg  
 961 gcggcgctgg agtggaagg caagccgtgg atcgacaagc cgatgatcgc cggccggaca  
 1021 tga

## 20 Seq. ID No.33

1 M K R A L I T G I T G Q D G S Y L A E L L L A K G Y E V H G  
 31 L I R R A S T F N T S R I D H L Y V D P H Q P G A R L F L H  
 61 Y G D L I D G T R L V T L L S T I E P D E V Y N L A A Q S H  
 91 V R V S F D E P V H T G D T T G M G S M R L L E A V R L S R  
 121 V H C R F Y Q A S S S E M F G A S P P P Q N E L T P F Y P R  
 15 151 S P Y G A A K V Y S Y W A T R N Y R E A Y G L F A V N G I L  
 181 F N H E S P R R G E T F V T R K I T R A V A R I K A G I Q S  
 211 E V Y M G N L D A V R D W G Y A P E Y V E G M W R M L Q T D  
 241 E P D D F V L A T G R G F T V R E F A R A A F E H A G L D W  
 30 271 Q Q Y V K F D Q R Y L R P T E V D S L I G D A T K A A E L L  
 301 G W R A S V H T D E L A R I M V D A D M A A L E C E G K P W  
 331 I D K P M I A G R T

## Seq. ID No.34

1 atgaggctgg ccgctcgcgc tcggaacatc ttgcgtcgca acggcatcga ggtgtcgcgc  
 35 61 tactttgccg aactggactg ggaacgcaat ttcttgccgc aactgcaatc gcatcgggtc  
 121 agtgccgtgc tcgatgtcgg ggccaattcg gggcagtagc ccaggggtct gcgcggcgcg  
 181 ggcttcgcgg gccgcategt ctgcttcgag ccgctgcccg ggccctttgc cgtcttcgag  
 241 cgcagcgccct ccacggaccc gttgtgggaa tgccggcgct gtgcgctggg cgatgtcgat  
 301 ggaaccatct cgatcaacgt cgcgggcaac gagggcgcca gcagttccgt cttgccgatg  
 40 361 ttgaaacgac atcaggacgc ctttccacca gccaaactac tgggcgccc accgggtgccg  
 421 atacatcgac tcgattccgt ggctgcagac gttctgcggc ccaacgatat tgcgttcttg  
 481 aagatcgacg ttcaaggatt cgagaagcag gtgatcgcg gtggcgattc aacgggtgca  
 541 gaccgatcgc tcggcatgca gctcagctg tctttccagc cgttgtacga ggggtggcatg  
 601 ctcatccggc aggcgctcga tctcgtggat tcgttgggtt ttacgctctc gggattgcaa  
 45 661 cccggtttca ccgacccccg caacggctga atgctgcagg ccgatggcat cttcttccgg  
 721 ggcagcgatt ga

- 55 -

## Seq. ID No.35

1 MRLARRARNILRRNGIEVSRYFAELDWERN  
 31 FLRQLQSHRVSAVLDVGANSQGQYARGLRGA  
 61 GFAGRIVSFEPPLPGPFAVLQRSASTDPLWE  
 5 91 CRRCALGDDVDGTISINVAGNEGASSSVLPM  
 121 LKRHQDAFPPANYVGAQRPVPIHRLDSVAAD  
 151 VLRPNIDIAFLKIDVQGFQKQVIAGGDSTVH  
 181 DRCVGMQLELSFQPLYEGGMLIREALDLVD  
 211 SLGFTLSGLQPGFTDPRNGRMLQADGIFFR  
 10 241 GSD

## Seq. ID No.36

1 gtgaaatcgt tgaaactcgc tcgttttcac gcgcgtagcg ccgccttcga ggtttcgcgc  
 61 cgctattctg agcgagacct gaagcaccag tttgtgaagc aactcaaadc gcgtcgggta  
 121 gatgtcggtt tcgatgtcgg cgccaactca ggacaatacg ccgcggcct ccgcgagca  
 15 181 gcatataagg gccgcattgt ctggttcgaa ccgctatccg gaccgtttac gatcttgaa  
 241 agcaaagcgt caacggatcc actttgggat tgccggcagc atgcgttggg cgattctgat  
 301 ggaacgggta cgatcaatat cgcaggaaac gccggtcaga gcagttcgt cttgcccatg  
 361 ctgaaaagtc atcagaacgc ttttccccgc gcaaaactatg tcggtaccca agaggcgccc  
 421 atacatcgac ttgattccgt ggcgccagaa tttctaggca tgaacgggtg cgcttttctc  
 20 481 aaggtcgacg ttcaaggctt tgaaaagcag gtgctcgccg ggggcaaadc aaccatagat  
 541 gaccattcgc tcggcatgca actcgaactg tccttctcgc cgttgtaga aggtggcatg  
 601 ctcatctcgc aagccctcga tctcgtgtat tccttgggct tcacgttgac gggattgctg  
 661 ccttggttca ttgatgcaa taatggtcga atgttgagg ccgacggcat ctttttcgcg  
 721 gaggacgatt ga

25

## Seq. ID No.37

1 MKSLKLARFIARSAAFEVSRRYSERDLKHQ  
 31 FVKQLKSRRVDVVFDFTVGANSQGQYAAGLR  
 61 RAAYKGRIVSFEPPLSGPFTILESASTDPL  
 91 WDCRQHALLGDSDDGTVTINIAAGNAGQSSSVL  
 121 PMLKSHQNAFPPANYVGTQEASIHRLDSVA  
 151 PEFLLGMNGVAFLKVDVQGFQKQVLAGGKST  
 181 IDDHCVGMQLELSFLPLYEGGMLIPEALDL  
 211 VYSLGFTLTGLLPCFIDANNRMLQADGIF  
 30 241 FRED D

- 56 -

## Seq. ID No.38

1 atggtgcaga cgaaacgata cgccggcttg accgcagcta acacaaagaa agtcgccatg  
61 gccgcaccaa tgttttcgat catcatcccc accttgaacg tggctgcggt attgcctgcc  
121 tgcttcgaca gcatcgcccg tcagacctgc ggtgacttcg agctgggtact ggtcgacggc  
5 181 ggctcgacgg acgaaaccct cgacatcgcc aacattttcg cccccaacct cggcgagcgg  
241 ttgatcatc atcgcgacac cgaccagggc gtctacgacg ccatgaaccg cggcgtggac  
301 ctggccaccg gaacgtggtt gctctttctg ggcgcggacg acagcctgta cgaggctgac  
361 accctggcgc ggggtggcgc cttcattggc gaacacgagc ccagcgatct ggtatatggc  
421 gacgtgatca tgcgctcaac caatttcgcg tggggtggcg ccttcgacct cgaccgtctg  
10 481 ttgttcaagc gcaacatctg ccatcaggcg atcttctacc gccgcggact cttcggcacc  
541 atcggctcct acaacctcgc ctaccgggtc ctggccgact gggacttcaa tattcgctgc  
601 ttttccaacc cagcgctcgt caccgctac atgcacgtgg tcgttgcaag ctacaacgaa  
661 ttcggcgggc tcagcaatac gatcgctgac aaggagtttt tgaagcggct gccgatgtcc  
721 acgagactcg gcataaggct ggcatagtt ctggtgcgca ggtggccaaa ggtgatcagc  
15 781 agggccatgg taatgcgcac cgtcatttct tggcggcgcc gacgttag

## Seq. ID No.39

1 M V Q T K R Y A G L T A A N T K K V A M A A P M F S I I I P  
31 T L N V A A V L P A C L D S I A R Q T C G D F E L V L V D G  
61 G S T D E T L D I A N I F A P N L G E R L I I H R D T D Q G  
20 91 V Y D A M N R G V D L A T G T W L L F L G A D D S L Y E A D  
121 T L A R V A A F I G E H E P S D L V Y G D V I M R S T N F R  
151 W G G A F D L D R L L F K R N I C H Q A I F Y R R G L F G T  
181 I G P Y N L R Y R V L A D W D F N I R C F S N P A L V T R Y  
211 M H V V V A S Y N E F G G L S N T I V D K E F L K R L P M S  
25 241 T R L G I R L V I V L V R R W P K V I S R A M V M R T V I S  
271 W R R R R

## Seq 40:

GATGCCGTGAGGAGGTAAAGCTGC

## Seq 41:

30 GATACGGCTCTTGAATCCTGCACG

CLAIMS

1. A polypeptide in substantially isolated form which comprises a sequence selected from the sequences of Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29, or a polypeptide substantially homologous thereto.
2. A polypeptide in substantially isolated form which comprises a sequence selected from the sequences of Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29.
3. A polypeptide which comprises a fragment of a polypeptide defined in claim 1 or 2, said fragment comprising at least 12 amino acids and an epitope.
4. A polynucleotide in substantially isolated form which encodes a polypeptide according to any one of claims 1 to 3.
5. A polynucleotide in substantially isolated form which is capable of selectively hybridizing to Seq.ID.No: 3 or 4 or a fragment thereof.
6. A polynucleotide fragment according to claim 5 which comprises a sequence selected from the sequences of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27, or a polynucleotide at least 90% homologous thereto.
7. A polynucleotide in substantially isolated form comprising a sequence selected from the sequences of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27.
8. A polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide as defined in any one of claims 4 to 7, optionally carrying a revealing label.

9. A recombinant vector carrying a polynucleotide as defined in any one of claims 4 to 7.

10. An antibody capable of binding a polypeptide or fragment thereof as defined in any one of claims 1 to 3.

11. An antibody capable of binding a polypeptide or fragment thereof wherein the polypeptide is a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or is a peptide substantially homologous thereto.

12. A test kit for detecting the presence or absence of a pathogenic mycobacterium in a sample which comprises a polynucleotide according to any one of claims 4 to 8, a polypeptide according to any one of claims 1 to 3, a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, or an antibody according to, any one of claims 10 or 11.

13. A method of detecting the presence or absence of antibodies in an animal or human, against a pathogenic mycobacteria in a sample which comprises:

- (a) providing a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which comprises an epitope;
- (b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said polypeptide is formed.

14. A method of detecting the presence or absence of a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the



sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto in a biological sample which method which comprises:

- (a) providing an antibody according to any one of claims 10 and 11;
- (b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said antibody is formed.

15. A method of detecting the presence or absence of cell mediated immune reactivity in an animal or human, to a polypeptide according to claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which method comprises

- (a) providing a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which comprises an epitope;
- (b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator or reaction to occur; and
- (c) detecting the presence of said cytokine or mediator or cellular response in the incubate.

16. A pharmaceutical composition comprising a polypeptide according to any one of claims 1 to 3 in a suitable carrier or diluent.

17. A Composition according to claim 16 or a composition comprising a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto,

for use in the treatment or prevention of diseases caused by mycobacteria.

18. A method of treating or preventing mycobacterial disease in an animal or human caused by mycobacteria which express a polypeptide according to claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which method comprises vaccinating or treating an animal or human with an effective amount of said polypeptide.

19. A method of treating or preventing mycobacterial diseases in animals or humans caused by mycobacteria containing the polynucleotide of Seq.ID.No: 3 or 4, which method comprises vaccinating or treating an animal or human with an effective amount of a polynucleotide according to claims 4 to 7, a vector according to claim 9 or a polynucleotide which encodes a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto.

20. A method according to claims 18 or 19 for increasing the in vivo susceptibility of mycobacteria to antimicrobial drugs.

21. A normally pathogenic mycobacterium, whose pathogenicity is mediated in all or in part by the presence or the expression of a polypeptide as defined in any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homologous thereto, which mycobacterium harbours an attenuating mutation in a gene encoding one of the said polypeptides.

22. A vaccine comprising a mycobacterium as claimed in claim 21.

23. A vaccine according to claim 22 wherein the mycobacteria is selected from *Mavs*, *Mptb* and *Mtb*.

00304-103440

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

HERMON-TAYLOR et al.

Serial No. To Be Assigned

Atty Ref.: 117-323

Filed: Concurrently Herewith

Group: Not Yet Assigned

For: NOVEL POLYNUCLEOTIDES AND  
POLYPEPTIDES IN PATHOGENIC  
MYCOBACTERIA AND THEIR USE AS  
DIAGNOSTICS, VACCINES AND TARGETS  
FOR CHEMOTHERAPY

Examiner: Not Yet Assigned

\* \* \* \* \*

November 6, 2000

Assistant Commissioner for Patents  
Washington, DC 20231

**SUBMISSION OF FORMAL DRAWINGS**

Sir:

Enclosed herewith is one (1) sheet of formal, inked drawings for the above-identified application.

Respectfully submitted,

**NIXON & VANDERHYE P.C.**

By: 

B.J. Sadoff

Reg. No. 36,663

BJS:jls  
1100 North Glebe Road, 8th Floor  
Arlington, VA 22201-4714  
Telephone: (703) 816-4000  
Facsimile: (703) 816-4100

Fig.1a)

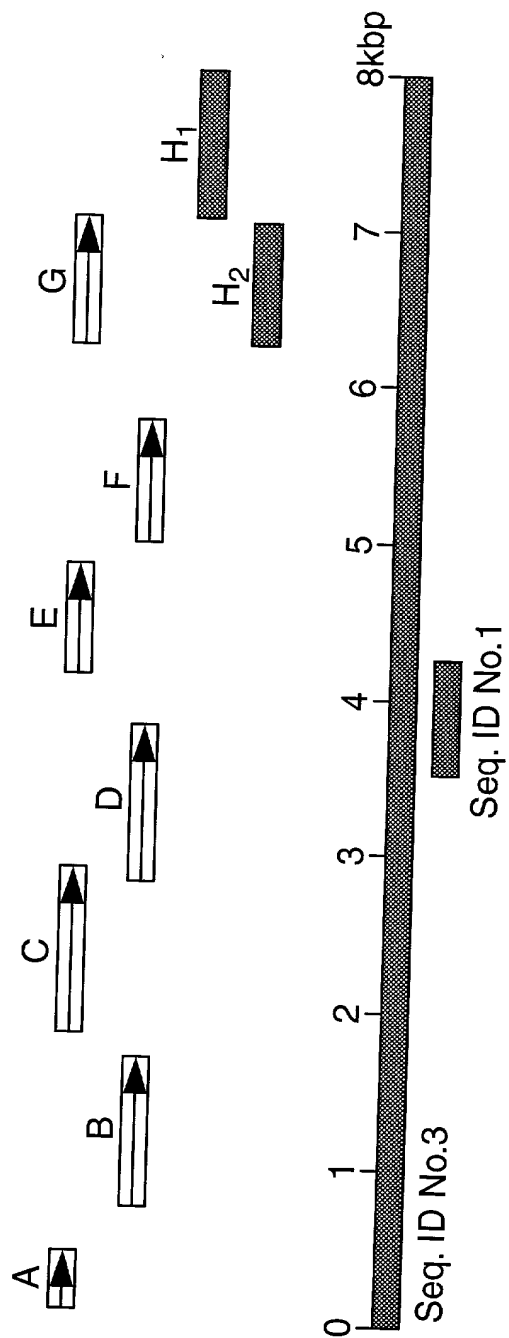
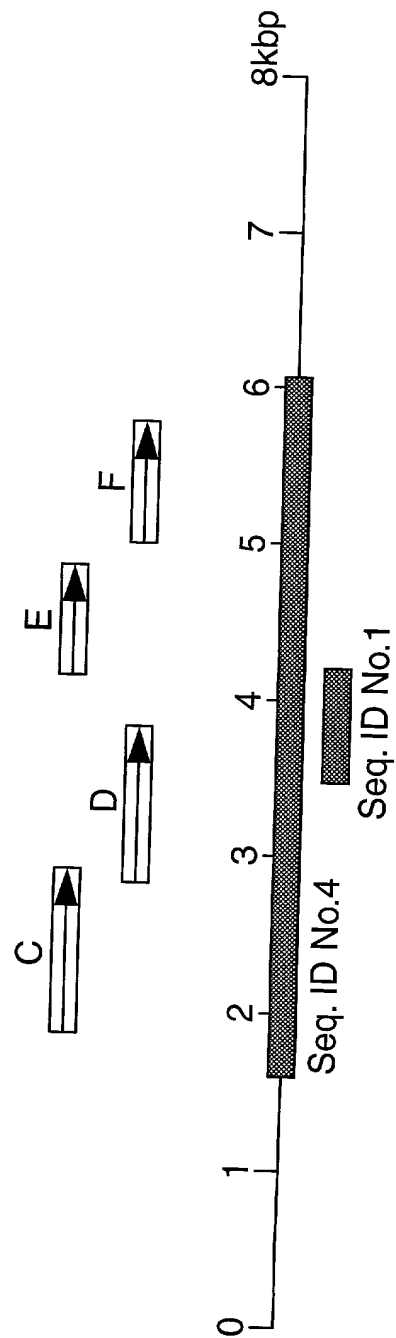


Fig.1b)



**RULE 63 (37 C.F.R. 1.63)**  
**DECLARATION AND POWER OF ATTORNEY**  
**FOR PATENT APPLICATION**  
**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

I, a below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name, and I believe in the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

**NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY**

the specification of which (check applicable box(es)):

☐ is attached hereto  
☐ was filed on 19 June 1998 as U.S. Application Serial No. (To Be Assigned) (Atty Dkt. No. 117-260)  
☒ was filed as PCT International application No. PCT/GB96/03221 on 23 December 1996  
and (if applicable to U.S. or PCT application) was amended on 22 December 1997

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with 37 C.F.R. 1.56. I hereby claim foreign priority benefits under 35 U.S.C. 119/365 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed or, if no priority is claimed, before the filing date of this application:

Priority Foreign Application(s):

Application Number	Country	Day/Month/Year Filed
9526178.0	Great Britain	21 December 1995

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

Application Number	Date/Month/Year Filed
--------------------	-----------------------

I hereby claim the benefit under 35 U.S.C. 120/365 of all prior United States and PCT International applications listed above or below and, insofar as the subject matter of each of the claims of this application is not disclosed in such prior applications in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose material information as defined in 37 C.F.R. 1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

Prior U.S./PCT Application(s):

Application Serial No.	Date/Month/Year Filed
PCT/GB96/03221	23 December 1996

Status: patented  
pending, abandoned

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon. And I hereby appoint NIXON & VANDERHYE P.C., 1100 North Glebe Rd., 8<sup>th</sup> Floor, Arlington, VA 22201-4714, telephone number (703) 816-4000 (to whom all communications are to be directed), and the following attorneys thereof (of the same address) individually and collectively my attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith and with the resulting patent: Arthur R. Crawford, 25327; Larry S. Nixon, 25640; Robert A. Vanderhye, 27076; James T. Hosmer, 30184; Robert W. Farris, 31352; Richard G. Besha, 22770; Mark E. Nusbaum, 32348; Michael J. Keenan, 32106; Bryan H. Davidson, 30251; Stanley C. Spooner, 27393; Leonard C. Mitchard, 29009; Duane M. Byers, 33363; Jeffrey H. Nelson, 30481; John R. Lastova, 33149; H. Warren Burnam, Jr. 29366; Thomas E. Byrne, 32205; Mary J. Wilson, 32955; J. Scott Davidson, 33489; Alan M. Kagen, 36178; William J. Griffin, 31260; Robert A. Molan, 29834; B. J. Sadoff, 36663; James D. Berquist, 34776; Updeep S. Gill, 37334.\*

1.	Inventor's Signature: <u>John Hermon-Taylor</u>	Date: <u>14<sup>th</sup> July 1998</u>	
	Inventor: <u>John</u> <u>HERMON-TAYLOR</u>		British
	(first) <u>MI</u> (last)		(citizenship)
	Residence: (city) <u>London</u>	(state/country) <u>United Kingdom</u>	
	Post Office Address: <u>St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom</u>		
	(Zip Code) <u>SW17 0RE</u>		
2.	Inventor's Signature: <u>Tim Doran</u>	Date: <u>5/8/1998</u>	
	Inventor: <u>Tim</u> <u>DORAN</u>		Australian
	(first) <u>MI</u> (last)		(citizenship)
	Residence: (city) <u>Whillington</u>	(state/country) <u>Australia</u>	
	Post Office Address: <u>1/8 Oxford Street, Whillington, Australia</u>		
	(Zip Code) <u>VIC 3219</u>		

FOR ADDITIONAL INVENTORS, check box ☒ and attach sheet with same information and signature and date for each.

RULE 63 (37 C.F.R. 1.63)  
DECLARATION AND POWER OF ATTORNEY  
FOR PATENT APPLICATION

Nixon &amp; Vanderhye P.C. (12/95)

Page 2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

3. Inventor's Signature: *Douglas* *MILLAR* Date: 14/8/98  
 Inventor: Douglas MILLAR  
 (first) (last)  
 Residence: (city) North Ryde (state/country) Australia  
 Post Office Address: Csiro Division of Biomolecular Engineering P.O. Box 184, North Ryde, Australia  
 (Zip Code) NSW 2113
4. Inventor's Signature: *Mark* *Tizard* Date: 5/8/98  
 Inventor: Mark TIZARD  
 (first) (last)  
 Residence: (city) London (state/country) United Kingdom  
 Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom  
 (Zip Code) SW17 0RE
5. Inventor's Signature: *M. Loughlin* Date: 24.7.98  
 Inventor: Mark LOUGHLIN  
 (first) (last)  
 Residence: (city) London (state/country) United Kingdom  
 Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom  
 (Zip Code) SW17 0RE
6. Inventor's Signature: *Nazira* *Sumar* Date: 14<sup>th</sup> July 1998  
 Inventor: Nazira SUMAR  
 (first) (last)  
 Residence: (city) London (state/country) United Kingdom  
 Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom  
 (Zip Code) SW17 0RE
7. Inventor's Signature: *John* *Ford* Date: 27/7/98  
 Inventor: John FORD  
 (first) (last)  
 Residence: (city) London (state/country) United Kingdom  
 Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom  
 (Zip Code) SW17 0RE
8. Inventor's Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
 Inventor: \_\_\_\_\_  
 (first) (last) (citizenship)  
 Residence: (city) \_\_\_\_\_ (state/country) \_\_\_\_\_  
 Post Office Address: \_\_\_\_\_  
 (Zip Code) \_\_\_\_\_
9. Inventor's Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
 Inventor: \_\_\_\_\_  
 (first) (last) (citizenship)  
 Residence: (city) \_\_\_\_\_ (state/country) \_\_\_\_\_  
 Post Office Address: \_\_\_\_\_  
 (Zip Code) \_\_\_\_\_
10. Inventor's Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
 Inventor: \_\_\_\_\_  
 (first) (last) (citizenship)  
 Residence: (city) \_\_\_\_\_ (state/country) \_\_\_\_\_  
 Post Office Address: \_\_\_\_\_  
 (Zip Code) \_\_\_\_\_
11. Inventor's Signature: \_\_\_\_\_ Date: \_\_\_\_\_  
 Inventor: \_\_\_\_\_  
 (first) (last) (citizenship)  
 Residence: (city) \_\_\_\_\_ (state/country) \_\_\_\_\_  
 Post Office Address: \_\_\_\_\_  
 (Zip Code) \_\_\_\_\_

## SEQUENCE LISTING

<110> Hermon-Taylor, John  
 Doran, Tim  
 Millar, Douglas  
 Tizard, Mark  
 Loughlin, Mark  
 Sumar, Nazira

<120> NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC  
 MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND  
 TARGETS FOR CHEMOTHERAPY

<130> 117-260

<140> 09/091,538

<141> 1998-06-19

<150> PCT/GB96/03221

<151> 1996-12-23

<150> GB 9526178.0

<151> 1995-12-21

<160> 41

<170> PatentIn Ver. 2.0

<210> 1

<211> 674

<212> DNA

<213> Mycobacterium

<400> 1

```

gatccaacta aacccgatgg aaccccgcg aaactattgg acgtctccgc gctacgcagt 60
tgggttggcg cccgcgaatc gcaactgaaag agggcatcga tgcaacgggtg tcgtggtacc 120
gcacaaatgc cgatgccgtg aggaggtaaa gctgcggggc ggccgatgtt atccctccgg 180
ccggacgggt agggcgacct gccatcgagt ggtacggcag tcgcctggcc ggcgaggcgc 240
atggcctatg tgagtatccc atagcctggc ttggctcgcc cctacgcatt atcagttgac 300
cgctttcgcg ccacgtcgca ggcttgcggc agcatcccgt tcaggtctcc tcatggtccg 360
gtgtggcacg accacgcaag ctcgaaccga ctcgtttccc aatttcgcat gctaatatcg 420
ctcgatggat tttttgcgca acgccggctt gatggctcgt aacgttagca ccgagatgct 480
gcgccactcc gaacgaaagc gcctattagt aaaccaagtc gaagcatacg gagtcaacgt 540
tgttattgat gtcggtgcta actccggcca gttcggtagc gctttgcgtc gtgcaggatt 600
caagagccgt atcgtttcct ttgaacctct ttcggggcca tttgcgcaac taacgcgcaa 660
gtcggcatcg gatc

```

674

<210> 2

<211> 674



<212> DNA

<213> Mycobacterium

<400> 2

gatccgatgc cgacttgcg gttagttgcg caaatggccc cgaaagaggt tcaaaggaaa 60  
 cgatacggct cttgaatcct gcacgacgca aagcgctacc gaactggccg gagttagcac 120  
 cgacatcaat aacaacgttg actccgtatg cttcgacttg gtttactaat aggcgctttc 180  
 gttcggagtg gcgcagcatc tcggtgctaa cgttacgagc catcaagccg gcgttgcgca 240  
 aaaaatccat cgagcgatat tagcatgcga aattgggaaa cgagtcgggt cgagcttgcg 300  
 tggtcgtgcc acaccggacc atgaggagac ctgaacggga tgctgccgca agcctgcgac 360  
 gtggcgcgaa agcgggtcaac tgataatgcg taggggagag ccaagccagg ctatgggata 420  
 ctcacatagg ccatgcgcct cgccggccag gcgactgccg taccactcga tggcaggtcg 480  
 ccctacccgt ccggccggag ggataacatc ggccggcccg cagctttacc tctcacggc 540  
 atcggcatTT gtgcggtacc acgacaccgt tgcacgatg ccctctttca gtgcgattcg 600  
 cgggcgccaa cccaactgcg tagcgcggag acgtccaata gtttgcgcg ggttccatcg 660  
 ggtttagttg gatc 674

<210> 3

<211> 7995

<212> DNA

<213> Mycobacterium

<400> 3

gaattctggg ttggagacga cgtcgaactc ctggtcggtc ttgcttcgaa tgatcgctgt 60  
 gatctggtcg gcggtgccga caggaaaccgt cgacttgctg acgatcacct tgtaccggtc 120  
 gatgtatgac ccaatgtcgt ccgcaaccga gaagacgtac gtcagggtcg ccgccccgct 180  
 ttcacccatg ggcgtcgga cggcgatgaa aatgacgtcc gcgtgctoga ttccgcgttg 240  
 ccggtcggtg gtgaagtcaa tcagcccggt ctcacggttc ctcgcaatca actcccaacc 300  
 cgggctcgaa aatcgggaca ctgcctgcga ggagcaaatc gatcttgcc tgatcgatat 360  
 cgacacagac gacatcggtg ccgctatccg cgagacaggc gccgtgacg aggcctacat 420  
 agcctgatcc gaccaccgaa attttcaaga tgacccttc aagtccccga tcggtcgacg 480  
 accatactgc cgcaactctg taccctccgt gggtaattcg catgtcgcgt tcgtaaggag 540  
 cagccagcga gtcggggacg ttcggtgaga gagtcgcagg actacgaggt tgccggtgcy 600  
 atacatcaca gtgttgcgtc tgcggcaac gatgcagcaa gaaccacgg ggcagccctg 660  
 aactgcgcgc atgaccggtc cttgtcctgg cacctttgat cggccaccgc ttccatgcga 720  
 acatgaccgg aatccatagc gcgtggtcaa gcagcgggga ggtagacgtc ggtgtcatct 780

gctccaaccg tgtcgggtgat aacgatttcg ctgaacgatc tcgagggatt gaaaagcacc 840  
 gtggagagcg ttcgcgcgca gcgctatggg gggcgaatcg agcacatcgt catcgacggg 900  
 ggatcgggcg acgccgtcgt ggagtatctg tccggcgatc ctggctttgc atattggcaa 960  
 tctcagcccc acaacgggag atatgacgcg atgaatcagg gcattgccca ttcgtcgggc 1020  
 gacctgttgt ggtttatgca ctccacggat cgtttctccg atccagatgc agtcgcttcc 1080  
 gtgggtggagg cgctctcggg gcatggacca gtacgtgatt tgtgggggta cgggaaaaac 1140  
 aaccttgtcg gactcgacgg caaaccactt tccctcggc cgtacggcta tatgccgttt 1200  
 aagatgcgga aatttctgct cggcgcgacg gttgcgcac aggcgacatt cttcggcgcg 1260  
 tcgctggtag ccaagttggg cggttacgat cttgattttg gactcgaggc ggaccagctg 1320  
 ttcactacc gtgccgcaact aatacggcct cccgtcacga tcgaccgctg ggtttgcgac 1380  
 ttcgatgtca cgggacctgg ttcaaccag cccatccgtg agcactatcg gacctgctg 1440  
 cggctctggg acctgcatgg cgactacccg ctgggtgggc gcagagtgtc gtgggcttac 1500  
 ttgcgtgtga aggagtactt gattcgggcc gacctggccg cattcaacgc ggtaaagttc 1560  
 ttgcgagcga agttcgccag agcttcgcgg aagcaaaatt catagaaacc aacttctact 1620  
 gcctgacctg agcagcgccg aggcgcgcag cgcgatcagt gcgacctgaa cggccagggtg 1680  
 gaaagcgcca ccgatcccgg caccgagtgc ctgacgcttc ggatcccttg caccacaacg 1740  
 agagtgagag cgccatgatg aggaaatata ggctgggagg agtcaacgcc ggagtgacaa 1800  
 aagtgagaac ccggtgaagc gagcgcttat aacagggatc acggggcagg atggttctta 1860  
 cctcgccgag ctactactga gcaagggata cgagggttcac gggctcgttc gtcgagcttc 1920  
 gacgtttaac acgtcgcgga tcgatcacct ctacgttgac ccacaccaac cgggcgcgcg 1980  
 cttgttcttg cactatgcag acctcactga cggcaccggg ttggtgacct tgcctcagcag 2040  
 tatcgaccgg gatgaggtct acaacctcgc agcgcagtcc catgtgcgcg tcagctttga 2100  
 cgagccagtg cataccggag acaccaccgg catgggatcg atccgacttc tggaagcagt 2160  
 ccgcctttct cgggtggact gccggttcta tcaggcttcc tcgtcggaga tgttcggcgc 2220  
 atctccgcca ccgcagaacg aatcgacgcc gttctatccc cgttcgccat acggcgcggc 2280  
 caaggtcttc tcgtactgga cgactcgcaa ctatcgagag gcgtacggat tattcgcagt 2340  
 gaatggcatc ttgttcaacc atgagtcgcc cgggcgcggc gagactttcg tgaccgaaa 2400  
 gatcacgcgt gccgtggcgc gcatccgagc tggcgctcaa tcggaggtct atatgggcaa 2460  
 cctcgatgcg atccgcgact ggggctacgc gccgaatat gtcgagggga tgtggaggat 2520  
 gttgcaagcg cctgaacctg atgactacgt cctggcgaca gggcgtggtt acaccgtacg 2580  
 tgagtctcgt caagctgctt ttgacctgt cgggctcgac tggcaaaagc gcgtcaagtt 2640

tgacgaccgc tatttgcgtc ccaccgaggt cgattcgcta gtaggagatg ccgacaaggc 2700  
 ggcccagtca ctcggtcgga aagcttcggt tcatactggt gaactcgcgc gcatcatggt 2760  
 ggacgcggac atcgccgcgt tggagtgcga tggcacacca tggatcgaca cgccgatgtt 2820  
 gcctggttg ggagagtaa gttgacgact acacctgggc ctctggaccg cgcaacgccc 2880  
 gtgtatatcg ccggtcatcg ggggctggtc ggctcagcgc tcgtacgtag atttgaggcc 2940  
 gaggggttca ccaatctcat tgtgcgatca cgcgatgaga ttgatctgac ggaccgagcc 3000  
 gcaacgtttg attttgtgtc tgagacaaga ccacaggtga tcatcgatgc ggccgcaagg 3060  
 gtgcggcgga tcatggcgaa taacacctat ccgcgggact tcttgtccga aaacctccga 3120  
 atccagacca atttgctcga cgcagctgtc gccgtgcgtg tgccgcggct ccttttccctc 3180  
 ggctcgatcat gcatctaccc gaagtacgct ccgcaacctt tccacgagag tgctttattg 3240  
 actggccctt tggagcccac caacgacgcg tatgcgatcg ccaagatcgc cggatatcctg 3300  
 caagttcagg cggttaggcg ccaatatggg ctggcgtgga tctctgcgat gccgaactaac 3360  
 ctctacggac ccggcgacaa cttctccccg tccgggtcgc atctcttgcc ggcgctcatc 3420  
 cgtcgatatg aggaagccaa agctggtggt gcagaagagg tgacgaattg ggggaccggt 3480  
 actccgcggc gcgaacttct gcatgtcgac gatctggcga gcgatgcct gttccttttg 3540  
 gaacatttgc atggtccgaa ccacgtcaac gtgggcaccg gcgtcgatca cagcattagc 3600  
 gagatcgag acatggtcgc tacagcgtg ggctacatcg gcgaaacacg ttgggatcca 3660  
 actaaaccgc atggaacccc gcgcaaaacta ttggacgtct ccgcgctacg cgagttgggt 3720  
 tggcgccgcg gaatcgcaact gaaagacggc atcgatgcaa cgggtgctgtg gtaccgcaca 3780  
 aatgccgatg ccgtgaggag gtaaagctgc gggtcggccg atgttatccc tccggccgga 3840  
 cgggtggggc gacctgccgt cgagtgttac ggcagtcgcc tggccggcga ggcgcgtggc 3900  
 ctatgggagt atccaatagc ctggcttggc tcgccctac gcattatcag ttgaccgctt 3960  
 tcgcgccagc tcgcaggctt gcggcagcat ccggttcagg tctcctcatg gtccggtgtg 4020  
 gcacgaccac gcaagctcga accgactcgt ttcccaattt cgcgatgctaa tatcgctcga 4080  
 tggatttttt gcgcaacgcc ggcttgatgg ctcgtaacgt tagtaccgag atgctgcgcc 4140  
 acttcgaacg aaagcgcta ttagtaaacc aattcaaagc atacggagtc aacgttggtta 4200  
 ttgatgtcgg tgctaactcc ggccagttcg gtagcgcttt gcgtcgtgca ggattcaaga 4260  
 gccgtatcgt ttcttttgaa cctctttcgg ggccatttgc gcaactaacg cgcaagtcgg 4320  
 catcggatcc actatgggag tgcaccagt atgccctagg cgacgccgat gagacgatta 4380  
 ccatcaatgt ggcaggcaat gcgggggcaa gtagttccgt gctgccgatg cttaaaagtc 4440

atcaagatgc ctttcctccc gccaattata ttggcaccga agacgttgca atacaccgcc 4500  
 ttgattcggg tgcacagaa tttctgaacc ctaccgatgt tactttcctg aagatcgacg 4560  
 tacaggggtt cgagaagcag gttatcacgg gcagtaagtc aacgcttaac gaaagctgcg 4620  
 tcggcatgca actcgaactt tcttttatcc cgttgtagca aggtgacatg ctgattcatg 4680  
 aagcgcttga acttgtctat tccctagggt tcagactgac gggtttggtg cccggcttta 4740  
 cggatccgcg caatggtcga atgcttcaag ctgacggcat tttcttccgt ggggacgatt 4800  
 gacataaatg ctccgtcggc accctgccgg tatccaaacg ggcgatctgg tgagccggcc 4860  
 tcccgggcac ctaatcgact atctaaattg aggcggccgc gacgtgcggc acgaacaggt 4920  
 ggccggctgc tagcggtaca cacgtcatga ctgcgccagt gttctcgata attatcccta 4980  
 ccttcaatgc agcggtagcg ctgcaagcct gcctcggaag catcgtcggg cagacctacc 5040  
 gggaagtgga agtggctcct gtcgacggcg gttcgaccga tcggaccctc gacatcgca 5100  
 acagtttccg cccggaactc ggctcgcgac tggctgttca cagcggggcc gatgatggcc 5160  
 cctacgacgc catgaaccgc ggcgtcggcg tggccacagg cgaatgggta ctttttttag 5220  
 gcgcgcagca caccctctac gaaccaacca cgttggccca ggtagccgct tttctcggcg 5280  
 accatgcggc aagccatctt gtctatggcg atgttgtagt gcgttcgacg aaaagccggc 5340  
 atgcgggacc tttcgacctc gaccgcctcc tatttgagac gaatttgtgc caccaatcga 5400  
 tcttttaccg ccgtgagctt ttcgacggca tcggccctta caacctgcgc taccgagtct 5460  
 gggcggactg ggacttcaat attcgtctgt tctccaaccc ggcgctgatt acccgctaca 5520  
 tggacgtcgt gatttccgaa tacaacgaca tgaccggctt cagcatgagg caggggactg 5580  
 ataaagagtt cagaaaacgg ctgccaatgt acttctgggt tgcaggggtg gagacttgca 5640  
 ggcgcatgct ggcgtttttg aaagacaagg agaatcgccg tctggccttg cgtacgcggt 5700  
 tgataagggg taaggccgtc tccaaagaac gaagcgcaga accgtagtcg cggatccaca 5760  
 ttggacttct ttaacgcgtt tgcgtcctga tccaccttcc aagcccgctc cgcgtaacgc 5820  
 ggcgcgca gaagtggcgc atatgcctc actgttctcg tgccagtgtg tggaaagcgt 5880  
 cgagcactct ggttcgcgtt cttgacgttc gcgccgctc ctagaggtag cgtgtcacgt 5940  
 gactgaagcc aatgagtgc actcggcgtc gcgaaagggt tcagtcgcgg ttgagcaaga 6000  
 caccgcaaga ctactggagt gcgtgcacaa gcgcctccag ctgcgcgctg aaagcggatg 6060  
 caaagggatt cgaagcttga gcaacatgcg aaggggagaa cggcctatga ggctgggaca 6120  
 ggttttcgat ccgcgcgcga atgcactgtc aatggccaag tagaagtccc cgctggtggc 6180  
 cagcagaagt cccactccg ctgcgggtgg ttggctaatt cttggcggtt cccttcttgt 6240  
 ggtcggcgtg gcgcacccg taggactcgc cggaggtgac gacgatgctg gcgtggtgca 6300

gcagccgatac gaggatgctg gcggcggtgg tgtgctcggg caggaatcgc cccattgtt 6360  
 cgaagggcca atgcgaggcg atggccaggg agcggcgctc gtagccggca gccacgagcc 6420  
 ggaacaacag ttgagtcccg gtgtcgtcga gcggggcgaa gccgatctcg tccaagatga 6480  
 ccagatccgc gcggagcagg gtgtcgatga tcttgccgac ggtgttgcg gccaggccgc 6540  
 ggtagaggac ctcgatcagg tcggcgggcg tgaagtagcg gactttgaat ccggcgtgga 6600  
 cggcagcgtg cccgcagccg atgagcaggt gacttttgcc cgtaccaggt gggccaatga 6660  
 ccgccaggtt ctgttgtgcc cgaatccatt ccaggctcga caggtagtcg aacgtggctg 6720  
 cggtgatcga cgatccggtg acgtcgaacc cgtcgagggt cttggtgacc gggaaggctg 6780  
 cggccttgag acggttggcg gtgttgagg catcgcgggc agcgatctcg gcctcaacca 6840  
 acgtccgcag gatctcctcc ggtgtccagc gttgcgtctt ggcgacttgc aacacctcg 6900  
 cggcgttgcg gcgcaccgtg gccagcttca accgccgcag cgcgcgtca aggtcagcag 6960  
 ccagcggcgc cgcgaggac ggtgccaccg gcttggcagc ggtggtcatg aggcgcctcc 7020  
 gtcgggtggtg ttgatcttgt aggcctccaa cgagcgggtc tcgacgggtg gcagatcgag 7080  
 cacgagtgcg tcgccggcg ggcggggtt tggggtgcg gcgccggcg ccaggatcga 7140  
 gcgcacgtcg gcagcgcga accggcgaaa cgcaaccgcc cggcgcagcg cgtcaatcaa 7200  
 agcctgttcg ccgtgggcg cgccaaggcc gagcagaatg tcgagttcgg atttcagtcg 7260  
 ggtgttgccg atcgagcag caaccagcag gaactgctgc gcttcggtc ccaatgcgca 7320  
 gaatcgttc tctgcttggg ttttcgggcg aggaccacgc gaggggtgcg gctcgggtcc 7380  
 gtcgtagtgt tcatcgagga tggacacctc acctgggctg acgagctcgt gctcggccac 7440  
 gatcacaccg gtcgcagggt ccaacaggat caggcgcca tgatcgacca ccaccgccac 7500  
 ggtggcaccg acgagccgct gaggcaccga gtaacgagct gagccgtaac ggatgcacga 7560  
 gaggcgctg accttacggc gcaccgacc cgagccgata gtcggccgca gcgagggcag 7620  
 ctccctcaag acggtgcgct cgtcaaccaa gcgatcgtt ggacggcg agatctccga 7680  
 gtggaccgtg gcattgacct cggcgacca tagttgcgc tggcggttga gggcacgtag 7740  
 gtcgacctgc tcaccggcta acgcagctt ggtcagcagc ggcaccgcaa ggtcgtcctg 7800  
 agcgtagcca cagaggttct ccacgatgcc cttcgattgc ggatccgcac cgtggcagaa 7860  
 gtccggaacg aagccatagt gggacgcgaa tcgcacataa tccggtgtt gaacaacaac 7920  
 attggcgacg acaccacctt tgaggcagcc catccggtcg gccaggatct tggcgggaac 7980  
 cccaccgatac gcctc 7995

<211> 4435

<212> DNA

<213> Mycobacterium

<400> 4

```

ttctactgcc tgacctgagc agcgccgagg cgcgacagcg gatcactgcg acctgaatgg 60
ccagggtgaa agcgccaccg atcccggcac cgagtgcctg acgattcgga tcccttgcac 120
cacaacgaga gtgagaccgc catgatgacg aaatatcggc tgggcggagt caacgccgga 180
gtgacaaaag tgagaacctg gtgaagcgag cgcttataac agggatcacg gggcaggatg 240
gttcctacct cgccgagcta ctactgagca agggatacga ggttcacggg ctcgttcgtc 300
gagcttcgac gtttaacacg tcgcggatcg atcacctcta cgttgacca caccaaccgg 360
gcgcgcgctt gttcttgac tatgcagacc tctactgacg caccgggttg gtgacctgc 420
tcagcagtat cgaccggat gaggtctaca acctcgacg gcagtcccat gtgcgcgtca 480
gctttgacga gccagtgcac accggagaca ccaccggcat gggatcgatc cgacttctgg 540
aagcagtcgg cctttctcgg gtggactgcc ggttctatca ggttcctcg tcggagatgt 600
tcggcgcatc tccgccaccg cagaacgaat cgacgcggtt ctatccccgt tcgccatacg 660
gcgcggccaa ggtcttctcg tactggacga ctcgcaacta tcgagaggcg tacggattat 720
tcgcagtga tggcatcttg ttcaaccatg agtcccccg gcgcggcgag actttcgtga 780
ccgaaagat cagcgtgcc gtggcgcgca tccgagctgg ctgccaatcg gaggtctata 840
tgggcaacct cgatgcgac cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt 900
ggaggatgtt gcaagcgctt gaacctgatg actacgtcct ggcgacaggg cgtggttaca 960
ccgtacgtga gttcgctcaa gctgcttttg accacgtcgg gctcgactgg caaaagcacg 1020
tcaagtttga cgaccgctat ttgcgcccc cagaggctga ttcgctagta ggagatgccg 1080
acagggcggc ccagtcactc ggctggaaag cttcggttca tactggtgaa ctgcgcgcga 1140
tcatggtgga cgcgacatc gccgcgtcgg agtgcgatgg cacaccatgg atcgacacgc 1200
cgatgttgcc tggttggggc ggagtaagtt gacgactaca cctgggcctc tggaccgcgc 1260
aacgcccgtg tatatcgccg gtcacgggg gctggtcggc tcagcgctcg tacgtagatt 1320
tgaggccgag gggttcacca atctcattgt gcgatcacgc gatgagattg atctgacgga 1380
ccgagccgca acgtttgatt ttgtgtctga gacaagacca caggatgatc tcgatgcggc 1440
cgacagggtc ggcgcatca tggcgaataa cacctatccc gggacttct tgtccgaaaa 1500
cctccgaatc cagaccaatt tgctcgacgc agctgtcgcc gtgcgtgtgc cgggctcct 1560
tttctcgggt tcgtcatgca tctaccgaa gtacgctccg caacctatcc acgagagtgc 1620
tttattgact ggcccttgg agcccaccaa cgacgcgtat gcgatcgcca agatcgccgg 1680

```

tatcctgcaa gttcaggcgg ttagggccca atatgggctg gcgtaggatct ctgcgatgcc 1740  
 gactaacctc tacggacccg gcgacaactt ctccccgtcc gggtagcatc tcttgccggc 1800  
 gctcatccgt cgatatgagg aagccaaagc tggtagtgca gaagaggtga cgaattgggg 1860  
 gaccggtact ccgcgggcgg aactttctgca tgtcgacgat ctggcgagcg catgcctgtt 1920  
 ccttttgaa catttcgatg gtccgaacca cgtcaacgtg ggcaccggcg tcgatcacag 1980  
 cattagcgag atcgagaca tggtagctac ggcggtgggc tacatcgggc aaacacgttg 2040  
 ggatccaact aaacccgatg gaaccccgcg caaactattg gacgtctccg cgctacgca 2100  
 gttgggttg gcgccgaa tcgcactgaa agacggcatc gatgcaacgg tgtcgtggtg 2160  
 ccgcacaaat gccgatgccg tgaggaggta aagctgcggg ccggccgatg ttatccctcc 2220  
 ggccggacgg gtagggcgac ctgccatcga gtggtacggc agtcgcctgg ccggcgaggc 2280  
 gcatggccta tgggagtatc ccatagcctg gcttggtcgc ccctacgca ttatcagttg 2340  
 accgctttcg cgccagctcg caggctcgcg gcagcatccc gttcaggtct cctcatggtc 2400  
 cgggtgtggca cgaccacgca agctcgaacc gactcgtttc ccaatttcgc atgctaatat 2460  
 cgctcgatgg attttttgcg caacgccggc ttgatggctc gtaacgttag caccgagatg 2520  
 ctgcgccact tcgaacgaaa gcgcctatta gtaaaccaat tcaaagcata cggagtcaac 2580  
 gttgttattg atgtcgtgct taactccggc cagttcggta gcgctttgcg tcgtgcagga 2640  
 ttcaagagcc gtatcgtttc ctttgaacct ctttcggggc catttgcgca actaacgcgc 2700  
 gagtcggcat cggatccact atgggagtgt caccagtatg ccctaggcga cgccgatgag 2760  
 acgattacca tcaatgtggc aggcaatgcg ggggcaagta gttccgtgct gccgatgctt 2820  
 aaaagtcac aagatgcctt tcctcccgcg aattatattg gcaccgaaga cgttgcaata 2880  
 caccgccttg attcggttgc atcagaattt ctgaacccta ccgatgttac ttctctgaag 2940  
 atcgacgtac agggtttcga gaagcagggt atcgcgggca gtaagtcaac gcttaacgaa 3000  
 agctgcgtcg gcatgcaact cgaactttct tttattccgt tgtagcaagg tgacatgctg 3060  
 attcatgaag cgcttgaact tgtctattcc ctaggtttca gactgacggg tttgttgccc 3120  
 ggatttacgg atccgcgcaa tggtagaatg cttcaagctg acggcatttt cttccgtggg 3180  
 gacgattgac ataaatgctt gcgtcggcac cctgccggtg tccaaacggg cgatctggtg 3240  
 agccggcctc ccgggcaoct aatcgactat ctaaattgag gcggccgca cgtgcggcac 3300  
 gaacaggtgg ccggctgcta gcgttacaca cgtcatgact gcgccagtgt tctcgataat 3360  
 tatccctacc ttcaatgcag cggtagcgt gcaagcctgc ctcggaagca tcgtcgggca 3420  
 gacctaccgg gaagtggaag tggctcttgt cgacggcggg tcgaccgatc ggacctcga 3480  
 catcgogaac agtttccgcc cggaactcgg ctcgcgactg gtcgttcaca gcgggcccga 3540

tgatggcccc tacgacgcca tgaaccgcgg cgtcggcgta gccacaggcg aatgggtact 3600  
 ttttttaggc gccgacgaca ccctctacga accaaccacg ttggcccagg tagccgcttt 3660  
 tctcggcgac catgcggcaa gccatcttgt ctatggcgat gttgtgatgc gttcgacgaa 3720  
 aagccggcat gccggacctt tcgacctga ccgcctccta tttgagacga atttgtgcca 3780  
 ccaatcgatc ttttaccgcc gtgagctttt cgacggcatc ggcccttaca acctgcgcta 3840  
 ccgagtctgg gcggactggg acttcaatat tcgctgcttc tccaaccggg cgctgattac 3900  
 ccgctacatg gacgtcgtga tttccgaata caacgacatg accggcttca gcatgaggca 3960  
 ggggactgat aaagagttca gaaaacggct gccaatgtac ttctgggttg cagggtggga 4020  
 gacttgacgg cgcatgctgg cgtttttgaa agacaaggag aatcgccgtc tggccttgcg 4080  
 tacgcgggtg ataagggtta aggccgtctc caaagaacga agcgcagAAC cgtagtcgcg 4140  
 gatccacatt ggacttcttt aacgcgtttg cgtcctgac cacccttcaa ccccgttccg 4200  
 cgtgacgcgg cgcgagaga gtggtcgcat atcgcgtcac tgttctcgtg ccagtgcctg 4260  
 gaaagcgtcg agcactctgg ttcgcgttct tgacgttcgc gcccgccct agaggtagcg 4320  
 tgtcacgtga ctgaagccaa tgagtgaac tcggcgtcgc gaaaggtttc agtcgcggtt 4380  
 gagcaagaca ccgcaagact actggagtgc gtgcacaagc gcctccagct cacgg 4435

<210> 5

<211> 378

<212> DNA

<213> Mycobacterium

<220>

<221> CDS

<222> (1)..(375)

<400> 5

atg atc gct gtg atc tgg tcg gcg gtg ccg aca gga acc gtc gac ttg	48
Met Ile Ala Val Ile Trp Ser Ala Val Pro Thr Gly Thr Val Asp Leu	
1 5 10 15	
tcg acg atc acc ttg tac cgg tcg atg tat gac cca atg tcg tcc gca	96
Ser Thr Ile Thr Leu Tyr Arg Ser Met Tyr Asp Pro Met Ser Ser Ala	
20 25 30	
acc gag aag acg tac gtc agg tcc gcc gcc ccg ctt tca ccc atg ggc	144
Thr Glu Lys Thr Tyr Val Arg Ser Ala Ala Pro Leu Ser Pro Met Gly	
35 40 45	
gtc ggg acg gcg atg aaa atg acg tcc gcg tgc tcg att ccg cgt tgc	192
Val Gly Thr Ala Met Lys Met Thr Ser Ala Cys Ser Ile Pro Arg Cys	
50 55 60	
cgg tcg gtg gtg aag tca atc agc ccg ttc tca cgg ttc ctc gca atc	240
Arg Ser Val Val Lys Ser Ile Ser Pro Phe Ser Arg Phe Leu Ala Ile	
65 70 75 80	



aac tcc caa ccc ggg ctc gaa aat cgg gac act gcc tgc gag gag caa 288  
 Asn Ser Gln Pro Gly Leu Glu Asn Arg Asp Thr Ala Cys Glu Glu Gln  
                     85                    90                    95

atc gat ctt ggc ctg atc gat atc gac aca gac gac atc gtt gcc gct 336  
 Ile Asp Leu Gly Leu Ile Asp Ile Asp Thr Asp Asp Ile Val Ala Ala  
                     100                    105                    110

atc cgc gag aca ggc gcc cgt gac gag gcc tac ata gcc tga 378  
 Ile Arg Glu Thr Gly Ala Arg Asp Glu Ala Tyr Ile Ala  
                     115                    120                    125

<210> 6

<211> 125

<212> PRT

<213> Mycobacterium

<400> 6

Met Ile Ala Val Ile Trp Ser Ala Val Pro Thr Gly Thr Val Asp Leu  
           1                    5                    10                    15

Ser Thr Ile Thr Leu Tyr Arg Ser Met Tyr Asp Pro Met Ser Ser Ala  
                     20                    25                    30

Thr Glu Lys Thr Tyr Val Arg Ser Ala Ala Pro Leu Ser Pro Met Gly  
                     35                    40                    45

Val Gly Thr Ala Met Lys Met Thr Ser Ala Cys Ser Ile Pro Arg Cys  
                     50                    55                    60

Arg Ser Val Val Lys Ser Ile Ser Pro Phe Ser Arg Phe Leu Ala Ile  
                     65                    70                    75                    80

Asn Ser Gln Pro Gly Leu Glu Asn Arg Asp Thr Ala Cys Glu Glu Gln  
                     85                    90                    95

Ile Asp Leu Gly Leu Ile Asp Ile Asp Thr Asp Asp Ile Val Ala Ala  
                     100                    105                    110

Ile Arg Glu Thr Gly Ala Arg Asp Glu Ala Tyr Ile Ala  
                     115                    120                    125

<210> 7

<211> 834

<212> DNA

<213> Mycobacterium

<220>

<221> CDS

<222> (1)..(831)

<400> 7

gtg tca tct gct cca acc gtg tcg gtg ata acg att tcg ctg aac gat 48  
 Val Ser Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Leu Asn Asp  
           1                    5                    10                    15

ctc gag gga ttg aaa agc acc gtg gag agc gtt cgc gcg cag cgc tat	96
Leu Glu Gly Leu Lys Ser Thr Val Glu Ser Val Arg Ala Gln Arg Tyr	
20 25 30	
ggg ggg cga atc gag cac atc gtc atc gac ggt gga tcg ggc gac gcc	144
Gly Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Ala	
35 40 45	
gtc gtg gag tat ctg tcc ggc gat cct ggc ttt gca tat tgg caa tct	192
Val Val Glu Tyr Leu Ser Gly Asp Pro Gly Phe Ala Tyr Trp Gln Ser	
50 55 60	
cag ccc gac aac ggg aga tat gac gcg atg aat cag ggc att gcc cat	240
Gln Pro Asp Asn Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala His	
65 70 75 80	
tcg tcg ggc gac ctg ttg tgg ttt atg cac tcc acg gat cgt ttc tcc	288
Ser Ser Gly Asp Leu Leu Trp Phe Met His Ser Thr Asp Arg Phe Ser	
85 90 95	
gat cca gat gca gtc gct tcc gtg gtg gag gcg ctc tcg ggg cat gga	336
Asp Pro Asp Ala Val Ala Ser Val Val Glu Ala Leu Ser Gly His Gly	
100 105 110	
cca gta cgt gat ttg tgg ggt tac ggg aaa aac aac ctt gtc gga ctc	384
Pro Val Arg Asp Leu Trp Gly Tyr Gly Lys Asn Asn Leu Val Gly Leu	
115 120 125	
gac ggc aaa cca ctt ttc cct cgg ccg tac ggc tat atg ccg ttt aag	432
Asp Gly Lys Pro Leu Phe Pro Arg Pro Tyr Gly Tyr Met Pro Phe Lys	
130 135 140	
atg cgg aaa ttt ctg ctc ggc gcg acg gtt gcg cat cag gcg aca ttc	480
Met Arg Lys Phe Leu Leu Gly Ala Thr Val Ala His Gln Ala Thr Phe	
145 150 155 160	
ttc ggc gcg tcg ctg gta gcc aag ttg ggc ggt tac gat ctt gat ttt	528
Phe Gly Ala Ser Leu Val Ala Lys Leu Gly Gly Tyr Asp Leu Asp Phe	
165 170 175	
gga ctc gag gcg gac cag ctg ttc atc tac cgt gcc gca cta ata cgg	576
Gly Leu Glu Ala Asp Gln Leu Phe Ile Tyr Arg Ala Ala Leu Ile Arg	
180 185 190	
cct ccc gtc acg atc gac cgc gtg gtt tgc gac ttc gat gtc acg gga	624
Pro Pro Val Thr Ile Asp Arg Val Val Cys Asp Phe Asp Val Thr Gly	
195 200 205	
cct ggt tca acc cag ccc atc cgt gag cac tat cgg acc ctg cgg cgg	672
Pro Gly Ser Thr Gln Pro Ile Arg Glu His Tyr Arg Thr Leu Arg Arg	
210 215 220	
ctc tgg gac ctg cat ggc gac tac ccg ctg ggt ggg cgc aga gtg tcg	720
Leu Trp Asp Leu His Gly Asp Tyr Pro Leu Gly Gly Arg Arg Val Ser	
225 230 235 240	
tgg gct tac ttg cgt gtg aag gag tac ttg att cgg gcc gac ctg gcc	768
Trp Ala Tyr Leu Arg Val Lys Glu Tyr Leu Ile Arg Ala Asp Leu Ala	
245 250 255	

gca ttc aac gcg gta aag ttc ttg cga gcg aag ttc gcc aga gct tcg 816  
 Ala Phe Asn Ala Val Lys Phe Leu Arg Ala Lys Phe Ala Arg Ala Ser  
                   260                  265                  270

cgg aag caa aat tca tag 834  
 Arg Lys Gln Asn Ser  
                   275

<210> 8  
 <211> 277  
 <212> PRT  
 <213> Mycobacterium

<400> 8  
 Val Ser Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Leu Asn Asp  
   1                  5                  10                  15  
 Leu Glu Gly Leu Lys Ser Thr Val Glu Ser Val Arg Ala Gln Arg Tyr  
                   20                  25                  30  
 Gly Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Ala  
                   35                  40                  45  
 Val Val Glu Tyr Leu Ser Gly Asp Pro Gly Phe Ala Tyr Trp Gln Ser  
   50                  55                  60  
 Gln Pro Asp Asn Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala His  
   65                  70                  75                  80  
 Ser Ser Gly Asp Leu Leu Trp Phe Met His Ser Thr Asp Arg Phe Ser  
                   85                  90                  95  
 Asp Pro Asp Ala Val Ala Ser Val Val Glu Ala Leu Ser Gly His Gly  
                   100                  105                  110  
 Pro Val Arg Asp Leu Trp Gly Tyr Gly Lys Asn Asn Leu Val Gly Leu  
                   115                  120                  125  
 Asp Gly Lys Pro Leu Phe Pro Arg Pro Tyr Gly Tyr Met Pro Phe Lys  
   130                  135                  140  
 Met Arg Lys Phe Leu Leu Gly Ala Thr Val Ala His Gln Ala Thr Phe  
   145                  150                  155                  160  
 Phe Gly Ala Ser Leu Val Ala Lys Leu Gly Gly Tyr Asp Leu Asp Phe  
                   165                  170                  175  
 Gly Leu Glu Ala Asp Gln Leu Phe Ile Tyr Arg Ala Ala Leu Ile Arg  
                   180                  185                  190  
 Pro Pro Val Thr Ile Asp Arg Val Val Cys Asp Phe Asp Val Thr Gly  
                   195                  200                  205  
 Pro Gly Ser Thr Gln Pro Ile Arg Glu His Tyr Arg Thr Leu Arg Arg  
   210                  215                  220  
 Leu Trp Asp Leu His Gly Asp Tyr Pro Leu Gly Gly Arg Arg Val Ser  
   225                  230                  235                  240

```
<220>  
<221> CDS  
<222> (1)..(1029)
```

<400> 9																			
gtg aag cga gcg ctt ata aca ggg atc acg ggg cag gat ggt tcc tac																	48		
Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr																			
1					5							10						15	
ctc gcc gag cta cta ctg agc aag gga tac gag gtt cac ggg ctc gtt																	96		
Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val																			
					20							25						30	
cgt cga gct tcg acg ttt aac acg tcg cgg atc gat cac ctc tac gtt																	144		
Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val																			
					35							40						45	
gac cca cac caa ccg ggc gcg cgc ttg ttc ttg cac tat gca gac ctc																	192		
Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu																			
					50							55						60	
act gac ggc acc ccg ttg gtg acc ctg ctc agc agt atc gac ccg gat																	240		
Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp																			
65					70							75						80	
gag gtc tac aac ctc gca gcg cag tcc cat gtg cgc gtc agc ttt gac																	288		
Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp																			
					85							90						95	
gag cca gtg cat acc gga gac acc acc ggc atg gga tcg atc cga ctt																	336		
Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu																			
					100							105						110	
ctg gaa gca gtc cgc ctt tct ccg gtg gac tgc cgg ttc tat cag gct																	384		
Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala																			
					115							120						125	
tcc tcg tcg gag atg ttc ggc gca tct ccg cca ccg cag aac gaa tcg																	432		
Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser																			
					130							135						140	
acg ccg ttc tat ccc cgt tcg cca tac ggc gcg gcc aag gtc ttc tcg																	480		
Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser																			
145					150							155						160	

tac tgg acg act cgc aac tat cga gag gcg tac gga tta ttc gca gtg 528  
 Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val  
 165 170 175

aat ggc atc ttg ttc aac cat gag tcc ccc cgg cgc ggc gag act ttc 576  
 Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe  
 180 185 190

gtg acc cga aag atc acg cgt gcc gtg gcg cgc atc cga gct ggc gtc 624  
 Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val  
 195 200 205

caa tcg gag gtc tat atg ggc aac ctc gat gcg atc cgc gac tgg ggc 672  
 Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly  
 210 215 220

tac gcg ccc gaa tat gtc gag ggg atg tgg agg atg ttg caa gcg cct 720  
 Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro  
 225 230 235 240

gaa cct gat gac tac gtc ctg gcg aca ggg cgt ggt tac acc gta cgt 768  
 Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg  
 245 250 255

gag ttc gct caa gct gct ttt gac cat gtc ggg ctc gac tgg caa aag 816  
 Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys  
 260 265 270

cgc gtc aag ttt gac gac cgc tat ttg cgt ccc acc gag gtc gat tcg 864  
 Arg Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285

cta gta gga gat gcc gac aag gcg gcc cag tca ctc ggc tgg aaa gct 912  
 Leu Val Gly Asp Ala Asp Lys Ala Ala Gln Ser Leu Gly Trp Lys Ala  
 290 295 300

tcg gtt cat act ggt gaa ctc gcg cgc atc atg gtg gac gcg gac atc 960  
 Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile  
 305 310 315 320

gcc gcg ttg gag tgc gat ggc aca cca tgg atc gac acg ccg atg ttg 1008  
 Ala Ala Leu Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu  
 325 330 335

cct ggt tgg ggc aga gta agt tga 1032  
 Pro Gly Trp Gly Arg Val Ser  
 340

&lt;210&gt; 10

&lt;211&gt; 343

&lt;212&gt; PRT

&lt;213&gt; Mycobacterium

&lt;400&gt; 10

Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr  
 1 5 10 15

Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val  
 20 25 30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val  
 35 40 45  
 Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu  
 50 55 60  
 Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp  
 65 70 75 80  
 Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp  
 85 90 95  
 Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu  
 100 105 110  
 Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala  
 115 120 125  
 Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser  
 130 135 140  
 Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser  
 145 150 155 160  
 Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val  
 165 170 175  
 Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe  
 180 185 190  
 Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val  
 195 200 205  
 Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly  
 210 215 220  
 Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro  
 225 230 235 240  
 Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg  
 245 250 255  
 Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys  
 260 265 270  
 Arg Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285  
 Leu Val Gly Asp Ala Asp Lys Ala Ala Gln Ser Leu Gly Trp Lys Ala  
 290 295 300  
 Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile  
 305 310 315 320  
 Ala Ala Leu Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu  
 325 330 335  
 Pro Gly Trp Gly Arg Val Ser  
 340

<210> 11  
 <211> 1032  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(1029)

<400> 11  
 gtg aag cga gcg ctt ata aca ggg atc acg ggg cag gat ggt tcc tac 48  
 Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr  
 1 5 10 15

ctc gcc gag cta cta ctg agc aag gga tac gag gtt cac ggg ctc gtt 96  
 Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val  
 20 25 30

cgt cga gct tcg acg ttt aac acg tcg cgg atc gat cac ctc tac gtt 144  
 Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val  
 35 40 45

gac cca cac caa ccg ggc gcg cgc ttg ttc ttg cac tat gca gac ctc 192  
 Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu  
 50 55 60

act gac ggc acc cgg ttg gtg acc ctg ctc agc agt atc gac ccg gat 240  
 Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp  
 65 70 75 80

gag gtc tac aac ctc gca gcg cag tcc cat gtg cgc gtc agc ttt gac 288  
 Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp  
 85 90 95

gag cca gtg cat acc gga gac acc acc ggc atg gga tcg atc cga ctt 336  
 Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu  
 100 105 110

ctg gaa gca gtc cgc ctt tct cgg gtg gac tgc cgg ttc tat cag gct 384  
 Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala  
 115 120 125

tcc tcg tcg gag atg ttc ggc gca tct ccg cca ccg cag aac gaa tcg 432  
 Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser  
 130 135 140

acg ccg ttc tat ccc cgt tcg cca tac ggc gcg gcc aag gtc ttc tcg 480  
 Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser  
 145 150 155 160

tac tgg acg act cgc aac tat cga gag gcg tac gga tta ttc gca gtg 528  
 Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val  
 165 170 175

aat ggc atc ttg ttc aac cat gag tcc ccc cgg cgc gcc gag act ttc 576  
 Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe  
 180 185 190

gtg acc cga aag atc acg cgt gcc gtg gcg cgc atc cga gct gcc gtc 624  
 Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val  
 195 200 205

caa tcg gag gtc tat atg ggc aac ctc gat gcg atc cgc gac tgg ggc 672  
 Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly  
 210 215 220

tac gcg ccc gaa tat gtc gag ggg atg tgg agg atg ttg caa gcg cct 720  
 Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro  
 225 230 235 240

gaa cct gat gac tac gtc ctg gcg aca ggg cgt ggt tac acc gta cgt 768  
 Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg  
 245 250 255

gag ttc gct caa gct gct ttt gac cac gtc ggg ctc gac tgg caa aag 816  
 Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys  
 260 265 270

cac gtc aag ttt gac gac cgc tat ttg cgc ccc acc gag gtc gat tcg 864  
 His Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285

cta gta gga gat gcc gac agg gcg gcc cag tca ctc ggc tgg aaa gct 912  
 Leu Val Gly Asp Ala Asp Arg Ala Ala Gln Ser Leu Gly Trp Lys Ala  
 290 295 300

tcg gtt cat act ggt gaa ctc gcg cgc atc atg gtg gac gcg gac atc 960  
 Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile  
 305 310 315 320

gcc gcg tcg gag tgc gat ggc aca cca tgg atc gac acg ccg atg ttg 1008  
 Ala Ala Ser Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu  
 325 330 335

cct ggt tgg ggc gga gta agt tga 1032  
 Pro Gly Trp Gly Gly Val Ser  
 340

<210> 12  
 <211> 343  
 <212> PRT  
 <213> Mycobacterium

<400> 12  
 Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr  
 1 5 10 15

Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val  
 20 25 30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val  
 35 40 45

Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu  
 50 55 60

Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp  
 65 70 75 80

Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp  
 85 90 95



Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu  
 100 105 110  
 Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala  
 115 120 125  
 Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser  
 130 135 140  
 Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser  
 145 150 155 160  
 Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val  
 165 170 175  
 Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe  
 180 185 190  
 Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val  
 195 200 205  
 Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly  
 210 215 220  
 Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro  
 225 230 235 240  
 Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg  
 245 250 255  
 Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys  
 260 265 270  
 His Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285  
 Leu Val Gly Asp Ala Asp Arg Ala Ala Gln Ser Leu Gly Trp Lys Ala  
 290 295 300  
 Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile  
 305 310 315 320  
 Ala Ala Ser Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu  
 325 330 335  
 Pro Gly Trp Gly Gly Val Ser  
 340

<210> 13  
 <211> 1020  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(1017)

&lt;400&gt; 13

gtg cga tgg cac acc atg gat cga cac gcc gat gtt gcc tgg ttg ggg	48
Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly	
1 5 10 15	
cag agt aag ttg acg act aca cct ggg cct ctg gac cgc gca acg ccc	96
Gln Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro	
20 25 30	
gtg tat atc gcc ggt cat cgg ggg ctg gtc ggc tca gcg ctc gta cgt	144
Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg	
35 40 45	
aga ttt gag gcc gag ggg ttc acc aat ctc att gtg cga tca cgc gat	192
Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp	
50 55 60	
gag att gat ctg acg gac cga gcc gca acg ttt gat ttt gtg tct gag	240
Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu	
65 70 75 80	
aca aga cca cag gtg atc atc gat gcg gcc gca cgg gtc ggc ggc atc	288
Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile	
85 90 95	
atg gcg aat aac acc tat ccc gcg gac ttc ttg tcc gaa aac ctc cga	336
Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg	
100 105 110	
atc cag acc aat ttg ctc gac gca gct gtc gcc gtg cgt gtg ccg cgg	384
Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg	
115 120 125	
ctc ctt ttc ctc ggt tcg tca tgc atc tac ccg aag tac gct ccg caa	432
Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln	
130 135 140	
cct atc cac gag agt gct tta ttg act ggc cct ttg gag ccc acc aac	480
Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn	
145 150 155 160	
gac gcg tat gcg atc gcc aag atc gcc ggt atc ctg caa gtt cag gcg	528
Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala	
165 170 175	
gtt agg cgc caa tat ggg ctg gcg tgg atc tct gcg atg ccg act aac	576
Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn	
180 185 190	
ctc tac gga ccc ggc gac aac ttc tcc ccg tcc ggg tcg cat ctc ttg	624
Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu	
195 200 205	
ccg gcg ctc atc cgt cga tat gag gaa gcc aaa gct ggt ggt gca gaa	672
Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu	
210 215 220	
gag gtg acg aat tgg ggg acc ggt act ccg cgg cgc gaa ctt ctg cat	720
Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His	
225 230 235 240	

gtc gac gat ctg gcg agc gca tgc ctg ttc ctt ttg gaa cat ttc gat 768  
 Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp  
 245 250 255

ggt ccg aac cac gtc aac gtg ggc acc ggc gtc gat cac agc att agc 816  
 Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser  
 260 265 270

gag atc gca gac atg gtc gct aca gcg gtg ggc tac atc ggc gaa aca 864  
 Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr  
 275 280 285

cgt tgg gat cca act aaa ccc gat gga acc ccg cgc aaa cta ttg gac 912  
 Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp  
 290 295 300

gtc tcc gcg cta cgc gag ttg ggt tgg cgc ccg cga atc gca ctg aaa 960  
 Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys  
 305 310 315 320

gac ggc atc gat gca acg gtg tgc tgg tac cgc aca aat gcc gat gcc 1008  
 Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala  
 325 330 335

gtg agg agg taa 1020  
 Val Arg Arg

<210> 14  
 <211> 339  
 <212> PRT  
 <213> Mycobacterium

<400> 14  
 Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly  
 1 5 10 15

Gln Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro  
 20 25 30

Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg  
 35 40 45

Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp  
 50 55 60

Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu  
 65 70 75 80

Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile  
 85 90 95

Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg  
 100 105 110

Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg  
 115 120 125

Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln  
 130 135 140

Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn  
 145 150 155 160  
 Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala  
 165 170 175  
 Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn  
 180 185 190  
 Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu  
 195 200 205  
 Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu  
 210 215 220  
 Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His  
 225 230 235 240  
 Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp  
 245 250 255  
 Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser  
 260 265 270  
 Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr  
 275 280 285  
 Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp  
 290 295 300  
 Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys  
 305 310 315 320  
 Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala  
 325 330 335  
 Val Arg Arg

<210> 15  
 <211> 1020  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(1017)

<400> 15  
 gtg cga tgg cac acc atg gat cga cac gcc gat gtt gcc tgg ttg ggg 48  
 Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly  
 1 5 10 15  
 cgg agt aag ttg acg act aca cct ggg cct ctg gac cgc gca acg ccc 96  
 Arg Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro  
 20 25 30  
 gtg tat atc gcc ggt cat cgg ggg ctg gtc ggc tca gcg ctc gta cgt 144  
 Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg  
 35 40 45

aga ttt gag gcc gag ggg ttc acc aat ctc att gtg cga tca cgc gat	192
Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp	
50 55 60	
gag att gat ctg acg gac cga gcc gca acg ttt gat ttt gtg tct gag	240
Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu	
65 70 75 80	
aca aga cca cag gtg atc atc gat gcg gcc gca cgg gtc ggc ggc atc	288
Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile	
85 90 95	
atg gcg aat aac acc tat ccc gcg gac ttc ttg tcc gaa aac ctc cga	336
Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg	
100 105 110	
atc cag acc aat ttg ctc gac gca gct gtc gcc gtg cgt gtg ccg cgg	384
Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg	
115 120 125	
ctc ctt ttc ctc ggt tcg tca tgc atc tac ccg aag tac gct ccg caa	432
Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln	
130 135 140	
cct atc cac gag agt gct tta ttg act ggc cct ttg gag ccc acc aac	480
Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn	
145 150 155 160	
gac gcg tat gcg atc gcc aag atc gcc ggt atc ctg caa gtt cag gcg	528
Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala	
165 170 175	
gtt agg cgc caa tat ggg ctg gcg tgg atc tot gcg atg ccg act aac	576
Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn	
180 185 190	
ctc tac gga ccc ggc gac aac ttc tcc ccg tcc ggg tcg cat ctc ttg	624
Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu	
195 200 205	
ccg gcg ctc atc cgt cga tat gag gaa gcc aaa gct ggt ggt gca gaa	672
Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu	
210 215 220	
gag gtg acg aat tgg ggg acc ggt act ccg cgg cgc gaa ctt ctg cat	720
Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His	
225 230 235 240	
gtc gac gat ctg gcg agc gca tgc ctg ttc ctt ttg gaa cat ttc gat	768
Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp	
245 250 255	
ggt ccg aac cac gtc aac gtg ggc acc ggc gtc gat cac agc att agc	816
Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser	
260 265 270	
gag atc gca gac atg gtc gct acg gcg gtg ggc tac atc ggc gaa aca	864
Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr	
275 280 285	

cgt tgg gat cca act aaa ccc gat gga acc cgg cgc aaa cta ttg gac 912  
 Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp  
 290 295 300  
  
 gtc tcc gcg cta cgc gag ttg ggt tgg cgc cgg cga atc gca ctg aaa 960  
 Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys  
 305 310 315 320  
  
 gac ggc atc gat gca acg gtg tcg tgg tac cgc aca aat gcc gat gcc 1008  
 Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala  
 325 330 335  
  
 gtg agg agg taa 1020  
 Val Arg Arg

<210> 16

<211> 339

<212> PRT

<213> Mycobacterium

<400> 16

Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly  
 1 5 10 15  
  
 Arg Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro  
 20 25 30  
  
 Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg  
 35 40 45  
  
 Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp  
 50 55 60  
  
 Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu  
 65 70 75 80  
  
 Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile  
 85 90 95  
  
 Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg  
 100 105 110  
  
 Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg  
 115 120 125  
  
 Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln  
 130 135 140  
  
 Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn  
 145 150 155 160  
  
 Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala  
 165 170 175  
  
 Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn  
 180 185 190  
  
 Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu  
 195 200 205

Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu  
 210 215 220  
 Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His  
 225 230 235 240  
 Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp  
 245 250 255  
 Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser  
 260 265 270  
 Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr  
 275 280 285  
 Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp  
 290 295 300  
 Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys  
 305 310 315 320  
 Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala  
 325 330 335  
 Val Arg Arg

<210> 17  
 <211> 723  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(720)

<400> 17  
 atg gat ttt ttg cgc aac gcc ggc ttg atg gct cgt aac gtt agt acc 48  
 Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr  
 1 5 10 15  
 gag atg ctg cgc cac ttc gaa cga aag cgc cta tta gta aac caa ttc 96  
 Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe  
 20 25 30  
 aaa gca tac gga gtc aac gtt gtt att gat gtc ggt gct aac tcc ggc 144  
 Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly  
 35 40 45  
 cag ttc ggt agc gct ttg cgt cgt gca gga ttc aag agc cgt atc gtt 192  
 Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val  
 50 55 60  
 tcc ttt gaa cct ctt tcg ggg cca ttt gcg caa cta acg cgc aag tcg 240  
 Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Lys Ser  
 65 70 75 80  
 gca tcg gat cca cta tgg gag tgt cac cag tat gcc cta ggc gac gcc 288  
 Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala  
 85 90 95

gat gag acg att acc atc aat gtg gca ggc aat gcg ggg gca agt agt 336  
 Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser  
 100 105 110

tcc gtg ctg ccg atg ctt aaa agt cat caa gat gcc ttt cct ccc gcg 384  
 Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala  
 115 120 125

aat tat att ggc acc gaa gac gtt gca ata cac cgc ctt gat tcg gtt 432  
 Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val  
 130 135 140

gca tca gaa ttt ctg aac cct acc gat gtt act ttc ctg aag atc gac 480  
 Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp  
 145 150 155 160

gta cag ggt ttc gag aag cag gtt atc acg ggc agt aag tca acg ctt 528  
 Val Gln Gly Phe Glu Lys Gln Val Ile Thr Gly Ser Lys Ser Thr Leu  
 165 170 175

aac gaa agc tgc gtc ggc atg caa ctc gaa ctt tct ttt att ccg ttg 576  
 Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu  
 180 185 190

tac gaa ggt gac atg ctg att cat gaa gcg ctt gaa ctt gtc tat tcc 624  
 Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser  
 195 200 205

cta ggt ttc aga ctg acg ggt ttg ttg ccc ggc ttt acg gat ccg cgc 672  
 Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg  
 210 215 220

aat ggt cga atg ctt caa gct gac ggc att ttc ttc cgt ggg gac gat 720  
 Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp  
 225 230 235 240

tga 723

<210> 18  
 <211> 240  
 <212> PRT  
 <213> Mycobacterium

<400> 18  
 Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr  
 1 5 10 15

Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe  
 20 25 30

Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly  
 35 40 45

Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val  
 50 55 60

Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Lys Ser  
 65 70 75 80



Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala  
85 90 95

Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser  
100 105 110

Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala  
115 120 125

Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val  
130 135 140

Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp  
145 150 155 160

Val Gln Gly Phe Glu Lys Gln Val Ile Thr Gly Ser Lys Ser Thr Leu  
165 170 175

Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu  
180 185 190

Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser  
195 200 205

Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg  
210 215 220

Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp  
225 230 235 240

<210> 19  
<211> 723  
<212> DNA  
<213> Mycobacterium

<220>  
<221> CDS  
<222> (1)..(720)

<400> 19  
atg gat ttt ttg cgc aac gcc ggc ttg atg gct cgt aac gtt agc acc 48  
Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr  
1 5 10 15

gag atg ctg cgc cac ttc gaa cga aag cgc cta tta gta aac caa ttc 96  
Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe  
20 25 30

aaa gca tac gga gtc aac gtt gtt att gat gtc ggt gct aac tcc ggc 144  
Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly  
35 40 45

cag ttc ggt agc gct ttg cgt cgt gca gga ttc aag agc cgt atc gtt 192  
Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val  
50 55 60

tcc ttt gaa cct ctt tcg ggg cca ttt gcg caa cta acg cgc gag tcg 240  
Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Glu Ser  
65 70 75 80

gca tog gat cca cta tgg gag tgt cac cag tat gcc cta ggc gac gcc 288  
 Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala  
                     85                    90                    95

gat gag acg att acc atc aat gtg gca ggc aat gcg ggg gca agt agt 336  
 Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser  
                     100                    105                    110

tcc gtg ctg ccg atg ctt aaa agt cat caa gat gcc ttt cct ccc gcg 384  
 Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala  
                     115                    120                    125

aat tat att ggc acc gaa gac gtt gca ata cac cgc ctt gat tog gtt 432  
 Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val  
                     130                    135                    140

gca tca gaa ttt ctg aac cct acc gat gtt act ttc ctg aag atc gac 480  
 Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp  
 145                    150                    155                    160

gta cag ggt ttc gag aag cag gtt atc gcg ggc agt aag tca acg ctt 528  
 Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Ser Lys Ser Thr Leu  
                     165                    170                    175

aac gaa agc tgc gtc ggc atg caa ctc gaa ctt tct ttt att ccg ttg 576  
 Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu  
                     180                    185                    190

tac gaa ggt gac atg ctg att cat gaa gcg ctt gaa ctt gtc tat tcc 624  
 Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser  
                     195                    200                    205

cta ggt ttc aga ctg acg ggt ttg ttg ccc gga ttt acg gat ccg cgc 672  
 Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg  
                     210                    215                    220

aat ggt cga atg ctt caa gct gac ggc att ttc ttc cgt ggg gac gat 720  
 Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp  
 225                    230                    235                    240

tga 723

<210> 20  
 <211> 240  
 <212> PRT  
 <213> Mycobacterium

<400> 20  
 Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr  
   1                    5                    10                    15

Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe  
                     20                    25                    30

Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly  
                     35                    40                    45

Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val  
                     50                    55                    60

Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Glu Ser  
 65 70 75 80  
 Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala  
 85 90 95  
 Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser  
 100 105 110  
 Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala  
 115 120 125  
 Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val  
 130 135 140  
 Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp  
 145 150 155 160  
 Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Ser Lys Ser Thr Leu  
 165 170 175  
 Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu  
 180 185 190  
 Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser  
 195 200 205  
 Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg  
 210 215 220  
 Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp  
 225 230 235 240

<210> 21  
 <211> 801  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(798)

<400> 21  
 atg act gcg cca gtg ttc tcg ata att atc cct acc ttc aat gca gcg 48  
 Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala  
 1 5 10 15  
 gtg acg ctg caa gcc tgc ctc gga agc atc gtc ggg cag acc tac cgg 96  
 Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg  
 20 25 30  
 gaa gtg gaa gtg gtc ctt gtc gac ggc ggt tcg acc gat cgg acc ctc 144  
 Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu  
 35 40 45  
 gac atc gcg aac agt ttc cgc ccg gaa ctc ggc tcg cga ctg gtc gtt 192  
 Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val  
 50 55 60

cac agc ggg ccc gat gat ggc ccc tac gac gcc atg aac cgc ggc gtc 240  
 His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val  
 65 70 75 80

ggc gtg gcc aca ggc gaa tgg gta ctt ttt tta ggc gcc gac gac acc 288  
 Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr  
 85 90 95

ctc tac gaa cca acc acg ttg gcc cag gta gcc gct ttt ctc ggc gac 336  
 Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp  
 100 105 110

cat gcg gca agc cat ctt gtc tat ggc gat gtt gtg atg cgt tcg acg 384  
 His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr  
 115 120 125

aaa agc cgg cat gcc gga cct ttc gac ctc gac cgc ctc cta ttt gag 432  
 Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu  
 130 135 140

acg aat ttg tgc cac caa tcg atc ttt tac cgc cgt gag ctt ttc gac 480  
 Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp  
 145 150 155 160

ggc atc ggc cct tac aac ctg cgc tac cga gtc tgg gcg gac tgg gac 528  
 Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp  
 165 170 175

ttc aat att cgc tgc ttc tcc aac ccg gcg ctg att acc cgc tac atg 576  
 Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met  
 180 185 190

gac gtc gtg att tcc gaa tac aac gac atg acc ggc ttc agc atg agg 624  
 Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg  
 195 200 205

cag ggg act gat aaa gag ttc aga aaa cgg ctg cca atg tac ttc tgg 672  
 Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp  
 210 215 220

gtt gca ggg tgg gag act tgc agg cgc atg ctg gcg ttt ttg aaa gac 720  
 Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp  
 225 230 235 240

aag gag aat cgc cgt ctg gcc ttg cgt acg cgg ttg ata agg gtt aag 768  
 Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys  
 245 250 255

gcc gtc tcc aaa gaa cga agc gca gaa ccg tag 801  
 Ala Val Ser Lys Glu Arg Ser Ala Glu Pro  
 260 265

<210> 22

<211> 266

<212> PRT

<213> Mycobacterium

<400> 22

Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala  
 1 5 10 15

Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg  
                   20                  25                  30  
 Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu  
                   35                  40                  45  
 Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val  
                   50                  55                  60  
 His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val  
                   65                  70                  75                  80  
 Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr  
                   85                  90                  95  
 Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp  
                   100                  105                  110  
 His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr  
                   115                  120                  125  
 Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu  
                   130                  135                  140  
 Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp  
                   145                  150                  155                  160  
 Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp  
                   165                  170                  175  
 Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met  
                   180                  185                  190  
 Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg  
                   195                  200                  205  
 Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp  
                   210                  215                  220  
 Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp  
                   225                  230                  235                  240  
 Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys  
                   245                  250                  255  
 Ala Val Ser Lys Glu Arg Ser Ala Glu Pro  
                   260                  265

<210> 23  
 <211> 801  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(798)

&lt;400&gt; 23

atg act gcg cca gtg ttc tcg ata att atc cct acc ttc aat gca gcg	48
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala	
1 5 10 15	
gtg acg ctg caa gcc tgc ctc gga agc atc gtc ggg cag acc tac cgg	96
Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg	
20 25 30	
gaa gtg gaa gtg gtc ctt gtc gac ggc ggt tcg acc gat cgg acc ctc	144
Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu	
35 40 45	
gac atc gcg aac agt ttc cgc ccg gaa ctc ggc tcg cga ctg gtc gtt	192
Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val	
50 55 60	
cac agc ggg ccc gat gat ggc ccc tac gac gcc atg aac cgc ggc gtc	240
His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val	
65 70 75 80	
ggc gta gcc aca ggc gaa tgg gta ctt ttt tta ggc gcc gac gac acc	288
Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr	
85 90 95	
ctc tac gaa cca acc acg ttg gcc cag gta gcc gct ttt ctc ggc gac	336
Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp	
100 105 110	
cat gcg gca agc cat ctt gtc tat ggc gat gtt gtg atg cgt tcg acg	384
His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr	
115 120 125	
aaa agc cgg cat gcc gga cct ttc gac ctc gac cgc ctc cta ttt gag	432
Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu	
130 135 140	
acg aa† ttg tgc cac caa tcg atc ttt tac cgc cgt gag ctt ttc gac	480
Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp	
145 150 155 160	
ggc atc ggc cct tac aac ctg cgc tac cga gtc tgg gcg gac tgg gac	528
Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp	
165 170 175	
ttc aat att cgc tgc ttc tcc aac ccg gcg ctg att acc cgc tac atg	576
Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met	
180 185 190	
gac gtc gtg att tcc gaa tac aac gac atg acc ggc ttc agc atg agg	624
Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg	
195 200 205	
cag ggg act gat aaa gag ttc aga aaa ccg ctg cca atg tac ttc tgg	672
Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp	
210 215 220	
gtt gca ggg tgg gag act tgc agg cgc atg ctg gcg ttt ttg aaa gac	720
Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp	
225 230 235 240	

aag gag aat cgc cgt ctg gcc ttg cgt acg cgg ttg ata agg gtt aag 768  
Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys  
245 250 255

```
gcc gtc tcc aaa gaa cga agc gca gaa ccg tag      801
Ala Val Ser Lys Glu Arg Ser Ala Glu Pro
          260          265
```

```
<210> 24
<211> 266
<212> PRT
<213> Mycobacterium
```

```

<400> 24
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala
  1          5          10          15
Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg
          20          25          30
Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu
          35          40          45
Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val
          50          55          60
His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val
  65          70          75          80
Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr
          85          90          95
Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp
          100          105          110
His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr
          115          120          125
Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu
          130          135          140
Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp
          145          150          155          160
Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp
          165          170          175
Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met
          180          185          190
Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg
          195          200          205
Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp
          210          215          220
Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp
          225          230          235          240

```

Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys  
 245 250 255

Ala Val Ser Lys Glu Arg Ser Ala Glu Pro  
 260 265

<210> 25  
 <211> 867  
 <212> DNA  
 <213> Mycobacterium

<220>  
 <221> CDS  
 <222> (1)..(864)

<400> 25  
 gtg gcc agc aga agt ccc cac tcc gct gcg ggt ggt tgg cta att ctt 48  
 Val Ala Ser Arg Ser Pro His Ser Ala Ala Gly Gly Trp Leu Ile Leu  
 1 5 10 15  
 ggc gcc tcc ctt ctt gtg gtc ggc gtg gcg cat ccg gta gga ctc gcc 96  
 Gly Gly Ser Leu Leu Val Val Gly Val Ala His Pro Val Gly Leu Ala  
 20 25 30  
 gga ggt gac gac gat gct ggc gtg gtg cag cag ccg atc gag gat gct 144  
 Gly Gly Asp Asp Asp Ala Gly Val Val Gln Gln Pro Ile Glu Asp Ala  
 35 40 45  
 ggc gcc ggt ggt gtg ctc ggg cag gaa tcg ccc cca ttg ttc gaa ggg 192  
 Gly Gly Gly Gly Val Leu Gly Gln Glu Ser Pro Pro Leu Phe Glu Gly  
 50 55 60  
 cca atg cga gcc gat gcc cag gga gcg gcg ctc gta gcc gcc agc cac 240  
 Pro Met Arg Gly Asp Gly Gln Gly Ala Ala Leu Val Ala Gly Ser His  
 65 70 75 80  
 gag ccg gaa caa cag ttg agt ccc ggt gtc gtc gag ccg gcc gaa gcc 288  
 Glu Pro Glu Gln Gln Leu Ser Pro Gly Val Val Glu Arg Gly Glu Ala  
 85 90 95  
 gat ctc gtc caa gat gac cag atc cgc gcg gag cag ggt gtc gat gat 336  
 Asp Leu Val Gln Asp Asp Gln Ile Arg Ala Glu Gln Gly Val Asp Asp  
 100 105 110  
 ctt gcc gac ggt gtt gtc ggc cag gcc gcg gta gag gac ctc gat cag 384  
 Leu Ala Asp Gly Val Val Gly Gln Ala Ala Val Glu Asp Leu Asp Gln  
 115 120 125  
 gtc gcc gcc ggt gaa gta gcg gac ttt gaa tcc gcc gtg gac gcc agc 432  
 Val Gly Gly Gly Glu Val Ala Asp Phe Glu Ser Gly Val Asp Gly Ser  
 130 135 140  
 gtg ccc gca gcc gat gag cag gtg act ttt gcc cgt acc agg tgg gcc 480  
 Val Pro Ala Ala Asp Glu Gln Val Thr Phe Ala Arg Thr Arg Trp Ala  
 145 150 155 160  
 aat gac cgc cag gtt ctg ttg tgc ccg aat cca ttc cag gct cga cag 528  
 Asn Asp Arg Gln Val Leu Leu Cys Pro Asn Pro Phe Gln Ala Arg Gln  
 165 170 175



gta gtc gaa cgt ggc tgc ggt gat cga cga tcc ggt gac gtc gaa ccc 576  
 Val Val Glu Arg Gly Cys Gly Asp Arg Arg Ser Gly Asp Val Glu Pro  
 180 185 190

gtc gag ggt ctt ggt gac cgg gaa ggc tgc ggc ctt gag acg gtt ggc 624  
 Val Glu Gly Leu Gly Asp Arg Glu Gly Cys Gly Leu Glu Thr Val Gly  
 195 200 205

ggt gtt gga ggc atc gcg ggc agc gat ctc ggc ctc aac caa cgt ccg 672  
 Gly Val Gly Gly Ile Ala Gly Ser Asp Leu Gly Leu Asn Gln Arg Pro  
 210 215 220

cag gat ctc ctc cgg tgt cca gcg ttg cgt ctt ggc gac ttg caa cac 720  
 Gln Asp Leu Leu Arg Cys Pro Ala Leu Arg Leu Gly Asp Leu Gln His  
 225 230 235 240

ctc ggc ggc gtt gcg gcg cac cgt ggc cag ctt caa ccg ccg cag cgc 768  
 Leu Gly Gly Val Ala Ala His Arg Gly Gln Leu Gln Pro Pro Gln Arg  
 245 250 255

cgc gtc aag gtc agc agc cag cgg tgc cgc cga gga cgg tgc cac cgg 816  
 Arg Val Lys Val Ser Ser Gln Arg Cys Arg Arg Gly Arg Cys His Arg  
 260 265 270

ctt ggc agc ggt ggt cat gag gcc gtc ccg tcg gtg gtg ttg atc ttg 864  
 Leu Gly Ser Gly Gly His Glu Ala Val Pro Ser Val Val Leu Ile Leu  
 275 280 285

tag 867

<210> 26  
 <211> 288  
 <212> PRT  
 <213> Mycobacterium

<400> 26  
 Val Ala Ser Arg Ser Pro His Ser Ala Ala Gly Gly Trp Leu Ile Leu  
 1 5 10 15

Gly Gly Ser Leu Leu Val Val Gly Val Ala His Pro Val Gly Leu Ala  
 20 25 30

Gly Gly Asp Asp Asp Ala Gly Val Val Gln Gln Pro Ile Glu Asp Ala  
 35 40 45

Gly Gly Gly Gly Val Leu Gly Gln Glu Ser Pro Pro Leu Phe Glu Gly  
 50 55 60

Pro Met Arg Gly Asp Gly Gln Gly Ala Ala Leu Val Ala Gly Ser His  
 65 70 75 80

Glu Pro Glu Gln Gln Leu Ser Pro Gly Val Val Glu Arg Gly Glu Ala  
 85 90 95

Asp Leu Val Gln Asp Asp Gln Ile Arg Ala Glu Gln Gly Val Asp Asp  
 100 105 110

Leu Ala Asp Gly Val Val Gly Gln Ala Ala Val Glu Asp Leu Asp Gln  
 115 120 125

Val Gly Gly Gly Glu Val Ala Asp Phe Glu Ser Gly Val Asp Gly Ser  
 130 135 140

Val Pro Ala Ala Asp Glu Gln Val Thr Phe Ala Arg Thr Arg Trp Ala  
 145 150 155 160

Asn Asp Arg Gln Val Leu Leu Cys Pro Asn Pro Phe Gln Ala Arg Gln  
 165 170 175

Val Val Glu Arg Gly Cys Gly Asp Arg Arg Ser Gly Asp Val Glu Pro  
 180 185 190

Val Glu Gly Leu Gly Asp Arg Glu Gly Cys Gly Leu Glu Thr Val Gly  
 195 200 205

Gly Val Gly Gly Ile Ala Gly Ser Asp Leu Gly Leu Asn Gln Arg Pro  
 210 215 220

Gln Asp Leu Leu Arg Cys Pro Ala Leu Arg Leu Gly Asp Leu Gln His  
 225 230 235 240

Leu Gly Gly Val Ala Ala His Arg Gly Gln Leu Gln Pro Pro Gln Arg  
 245 250 255

Arg Val Lys Val Ser Ser Gln Arg Cys Arg Arg Gly Arg Cys His Arg  
 260 265 270

Leu Gly Ser Gly Gly His Glu Ala Val Pro Ser Val Val Leu Ile Leu  
 275 280 285

&lt;210&gt; 27

&lt;211&gt; 1739

&lt;212&gt; DNA

&lt;213&gt; Mycobacterium

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(945)

&lt;400&gt; 27

atg ggc tgc ctc aaa ggt ggt gtc gtc gcc aat gtt gtt gtt cca aca	48
Met Gly Cys Leu Lys Gly Gly Val Val Ala Asn Val Val Val Pro Thr	
1 5 10 15	
ccg gat tat gtg cga ttc gcg tcc cac tat ggc ttc gtt ccg gac ttc	96
Pro Asp Tyr Val Arg Phe Ala Ser His Tyr Gly Phe Val Pro Asp Phe	
20 25 30	
tgc cac ggt gcg gat ccg caa tcg aag ggc atc gtg gag aac ctc tgt	144
Cys His Gly Ala Asp Pro Gln Ser Lys Gly Ile Val Glu Asn Leu Cys	
35 40 45	
ggc tac gct cag gac gac ctt gcg gtg ccg ctg ctg acc gaa gct gcg	192
Gly Tyr Ala Gln Asp Asp Leu Ala Val Pro Leu Leu Thr Glu Ala Ala	
50 55 60	
tta gcc ggt gag cag gtc gac cta cgt gcc ctc aac gcc cag gcg caa	240
Leu Ala Gly Glu Gln Val Asp Leu Arg Ala Leu Asn Ala Gln Ala Gln	
65 70 75 80	

cta tgg tgc gcc gag gtc aat gcc acg gtc cac tcg gag atc tgc gcc	288
Leu Trp Cys Ala Glu Val Asn Ala Thr Val His Ser Glu Ile Cys Ala	
85 90 95	
gtg ccc aac gat cgc ttg gtt gac gag cgc acc gtc ttg agg gag ctg	336
Val Pro Asn Asp Arg Leu Val Asp Glu Arg Thr Val Leu Arg Glu Leu	
100 105 110	
ccc tcg ctg cgg ccg acg atc ggc tcg ggg tcg gtg cgc cgt aag gtc	384
Pro Ser Leu Arg Pro Thr Ile Gly Ser Gly Ser Val Arg Arg Lys Val	
115 120 125	
gac ggc ctc tcg tgc atc cgt tac ggc tca gct cgt tac tcg gtg cct	432
Asp Gly Leu Ser Cys Ile Arg Tyr Gly Ser Ala Arg Tyr Ser Val Pro	
130 135 140	
cag cgg ctc gtc ggt gcc acc gtg gcg gtg gtg gtc gat cat ggc gcc	480
Gln Arg Leu Val Gly Ala Thr Val Ala Val Val Val Asp His Gly Ala	
145 150 155 160	
ctg atc ctg ttg gaa cct gcg acc ggt gtg atc gtg gcc gag cac gag	528
Leu Ile Leu Leu Glu Pro Ala Thr Gly Val Ile Val Ala Glu His Glu	
165 170 175	
ctc gtc agc cca ggt gag gtg tcc atc ctc gat gaa cac tac gac gga	576
Leu Val Ser Pro Gly Glu Val Ser Ile Leu Asp Glu His Tyr Asp Gly	
180 185 190	
ccc aga ccc gca ccc tcg cgt ggt cct cgc ccg aaa acc caa gca gag	624
Pro Arg Pro Ala Pro Ser Arg Gly Pro Arg Pro Lys Thr Gln Ala Glu	
195 200 205	
aaa cga ttc tgc gca ttg gga acc gaa gcg cag cag ttc ctc gtc ggt	672
Lys Arg Phe Cys Ala Leu Thr Glu Ala Gln Gln Phe Leu Val Gly	
210 215 220	
gct gct gcg atc ggc aac acc cga ctg aaa tcc gaa ctc gac att ctg	720
Ala Ala Ala Ile Gly Asn Thr Arg Leu Lys Ser Glu Leu Asp Ile Leu	
225 230 235 240	
ctc ggc ctt ggc gcc gcc cac ggc gaa cag gct ttg att gac gcg ctg	768
Leu Gly Leu Gly Ala Ala His Gly Glu Gln Ala Leu Ile Asp Ala Leu	
245 250 255	
cgc cgg gcg gtt gcg ttt cgc ccg ttc cgc gct gcc gac gtg cgc tcg	816
Arg Arg Ala Val Ala Phe Arg Arg Phe Arg Ala Ala Asp Val Arg Ser	
260 265 270	
atc ctg gcc gcc ggc gcc ggc acc cca caa ccc cgc ccc gcc ggc gac	864
Ile Leu Ala Ala Gly Ala Gly Thr Pro Gln Pro Arg Pro Ala Gly Asp	
275 280 285	
gca ctc gtg ctc gat ctg ccc acc gtc gag acc cgc tcg ttg gag gcc	912
Ala Leu Val Leu Asp Leu Pro Thr Val Glu Thr Arg Ser Leu Glu Ala	
290 295 300	
tac aag atc aac acc acc gac ggg acg gcc tca tgaccaccgc tgccaagccg	965
Tyr Lys Ile Asn Thr Thr Asp Gly Thr Ala Ser	
305 310 315	
gtggcaccgt cctcggcggc accgctggct gctgaccttg acgcggcgct gcggcggttg	1025

aagctggcca cggtagcgccg caacgccgcc gaggtgttgc aagtcgcca gacgcaacgc 1085  
 tggacaccgg aggagatcct ggggacgttg gttgaggccg agatcgctgc ccgcgatgcc 1145  
 tccaacaccg ccaaccgtct caaggccgca gccttcccgg tcaccaagac cctcgacggg 1205  
 ttcgacgtca ccggatcgtc gatcaccgca gccacgttcg actacctgtc gagcctggaa 1265  
 tggattcggg cacaacagaa cctggcggtc attggcccac ctggtacggg caaaagtcac 1325  
 ctgctcatcg gctgcgggca cgctgccgtc caccgccgat tcaaagtcgg ctacttcacc 1385  
 gccgccgacc tgatcgaggt cctctaccgc ggcttgccg acaacaccgt cggcaagatc 1445  
 atcgacaccc tgctccgcgc ggatctggtc atcttgacg agatcggctt cgcctcgctc 1505  
 gacgacaccg ggactcaact gttgttcggg ctctgggtg ccggctacga gcgccgctcc 1565  
 ctggccatcg cctcgattg gcccttcgaa caatgggggc gattcctgcc cgagcacacc 1625  
 accgccgcca gcacctcga tcggctgctg caccacgcca gcacgtcgt cacctccggc 1685  
 gagtctacc ggatgcgcca cgcgcaccac aagaaggag ccgccaagaa ttag 1739

<210> 28

<211> 315

<212> PRT

<213> Mycobacterium

<400> 28

Met Gly Cys Leu Lys Gly Gly Val Val Ala Asn Val Val Val Pro Thr  
           1                  5                  10                  15

Pro Asp Tyr Val Arg Phe Ala Ser His Tyr Gly Phe Val Pro Asp Phe  
                   20                  25                  30

Cys His Gly Ala Asp Pro Gln Ser Lys Gly Ile Val Glu Asn Leu Cys  
           35                  40                  45

Gly Tyr Ala Gln Asp Asp Leu Ala Val Pro Leu Leu Thr Glu Ala Ala  
           50                  55                  60

Leu Ala Gly Glu Gln Val Asp Leu Arg Ala Leu Asn Ala Gln Ala Gln  
           65                  70                  75                  80

Leu Trp Cys Ala Glu Val Asn Ala Thr Val His Ser Glu Ile Cys Ala  
                   85                  90                  95

Val Pro Asn Asp Arg Leu Val Asp Glu Arg Thr Val Leu Arg Glu Leu  
                   100                  105                  110

Pro Ser Leu Arg Pro Thr Ile Gly Ser Gly Ser Val Arg Arg Lys Val  
           115                  120                  125

Asp Gly Leu Ser Cys Ile Arg Tyr Gly Ser Ala Arg Tyr Ser Val Pro  
           130                  135                  140

Gln Arg Leu Val Gly Ala Thr Val Ala Val Val Val Asp His Gly Ala  
           145                  150                  155                  160

Leu Ile Leu Leu Glu Pro Ala Thr Gly Val Ile Val Ala Glu His Glu  
 165 170 175  
 Leu Val Ser Pro Gly Glu Val Ser Ile Leu Asp Glu His Tyr Asp Gly  
 180 185 190  
 Pro Arg Pro Ala Pro Ser Arg Gly Pro Arg Pro Lys Thr Gln Ala Glu  
 195 200 205  
 Lys Arg Phe Cys Ala Leu Gly Thr Glu Ala Gln Gln Phe Leu Val Gly  
 210 215 220  
 Ala Ala Ala Ile Gly Asn Thr Arg Leu Lys Ser Glu Leu Asp Ile Leu  
 225 230 235 240  
 Leu Gly Leu Gly Ala Ala His Gly Glu Gln Ala Leu Ile Asp Ala Leu  
 245 250 255  
 Arg Arg Ala Val Ala Phe Arg Arg Phe Arg Ala Ala Asp Val Arg Ser  
 260 265 270  
 Ile Leu Ala Ala Gly Ala Gly Thr Pro Gln Pro Arg Pro Ala Gly Asp  
 275 280 285  
 Ala Leu Val Leu Asp Leu Pro Thr Val Glu Thr Arg Ser Leu Glu Ala  
 290 295 300  
 Tyr Lys Ile Asn Thr Thr Asp Gly Thr Ala Ser  
 305 310 315

<210> 29

<211> 264

<212> PRT

<213> Mycobacterium

<220>

<221> DOMAIN

<222> (1)..(264)

<223> amino acid sequence is encoded by nucleotides  
945-1736 of SEQ ID NO:27

<400> 29

Met Thr Thr Ala Ala Lys Pro Val Ala Pro Ser Ser Ala Ala Pro Leu  
 1 5 10 15  
 Ala Ala Asp Leu Asp Ala Ala Leu Arg Leu Lys Leu Ala Thr Val  
 20 25 30  
 Arg Arg Asn Ala Ala Glu Val Leu Gln Val Ala Lys Thr Gln Arg Trp  
 35 40 45  
 Thr Pro Glu Glu Ile Leu Arg Thr Leu Val Glu Ala Glu Ile Ala Ala  
 50 55 60  
 Arg Asp Ala Ser Asn Thr Ala Asn Arg Leu Lys Ala Ala Ala Phe Pro  
 65 70 75 80  
 Val Thr Lys Thr Leu Asp Gly Phe Asp Val Thr Gly Ser Ser Ile Thr  
 85 90 95

Ala Ala Thr Phe Asp Tyr Leu Ser Ser Leu Glu Trp Ile Arg Ala Gln  
 100 105 110

Gln Asn Leu Ala Val Ile Gly Pro Pro Gly Thr Gly Lys Ser His Leu  
 115 120 125

Leu Ile Gly Cys Gly His Ala Ala Val His Ala Gly Phe Lys Val Arg  
 130 135 140

Tyr Phe Thr Ala Ala Asp Leu Ile Glu Val Leu Tyr Arg Gly Leu Ala  
 145 150 155 160

Asp Asn Thr Val Gly Lys Ile Ile Asp Thr Leu Leu Arg Ala Asp Leu  
 165 170 175

Val Ile Leu Asp Glu Ile Gly Phe Ala Pro Leu Asp Asp Thr Gly Thr  
 180 185 190

Gln Leu Leu Phe Arg Leu Val Ala Ala Gly Tyr Glu Arg Arg Ser Leu  
 195 200 205

Ala Ile Ala Ser His Trp Pro Phe Glu Gln Trp Gly Arg Phe Leu Pro  
 210 215 220

Glu His Thr Thr Ala Ala Ser Ile Leu Asp Arg Leu Leu His His Ala  
 225 230 235 240

Ser Ile Val Val Thr Ser Gly Glu Ser Tyr Arg Met Arg His Ala Asp  
 245 250 255

His Lys Lys Gly Ala Ala Lys Asn  
 260

<210> 30

<211> 789

<212> DNA

<213> Mycobacterium

<220>

<221> CDS

<222> (1)..(786)

<400> 30

gtg	acg	tct	gct	ccg	acc	gtc	tcg	gtg	ata	acg	atc	tcg	ttc	aac	gac	48
Met	Thr	Ser	Ala	Pro	Thr	Val	Ser	Val	Ile	Thr	Ile	Ser	Phe	Asn	Asp	
1				5					10					15		
ctc	gac	ggg	ttg	cag	cgc	acg	gtg	aaa	agt	gtg	cgg	gcg	caa	cgc	tac	96
Leu	Asp	Gly	Leu	Gln	Arg	Thr	Val	Lys	Ser	Val	Arg	Ala	Gln	Arg	Tyr	
			20					25					30			
cgg	gga	cgc	atc	gag	cac	atc	gta	atc	gac	ggg	ggc	agc	ggc	gac	gac	144
Arg	Gly	Arg	Ile	Glu	His	Ile	Val	Ile	Asp	Gly	Gly	Ser	Gly	Asp	Asp	
			35				40					45				
gtg	gtg	gca	tac	ctg	tcc	ggg	tgt	gaa	cca	ggc	ttc	gcg	tat	tgg	cag	192
Val	Val	Ala	Tyr	Leu	Ser	Gly	Cys	Glu	Pro	Gly	Phe	Ala	Tyr	Trp	Gln	
			50			55					60					

```

tcc gag ccc gac ggc ggg cgg tac gac gcg atg aac cag ggc atc gcg 240
Ser Glu Pro Asp Gly Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala
65 70 75 80

cac gca tcg ggt gat ctg ttg tgg ttc ttg cac tcc gcc gat cgt ttt 288
His Ala Ser Gly Asp Leu Leu Trp Phe Leu His Ser Ala Asp Arg Phe
85 90 95

tcc ggg ccc gac gtg gta gcc cag gcc gtg gag gcg cta tcc ggc aag 336
Ser Gly Pro Asp Val Val Ala Gln Ala Val Glu Ala Leu Ser Gly Lys
100 105 110

gga ccg gtg tcc gaa ttg tgg ggc ttc ggg atg gat cgt ctc gtc ggg 384
Gly Pro Val Ser Glu Leu Trp Gly Phe Gly Met Asp Arg Leu Val Gly
115 120 125

ctc gat ccg gtg cgc ggc ccg ata cct ttc agc ctg cgc aaa ttc ctg 432
Leu Asp Arg Val Arg Gly Pro Ile Pro Phe Ser Leu Arg Lys Phe Leu
130 135 140

gcc ggc aag cag gtt gtt ccg cat caa gca tcg ttc ttc gga tca tcg 480
Ala Gly Lys Gln Val Val Pro His Gln Ala Ser Phe Phe Gly Ser Ser
145 150 155 160

ctg gtg gcc aag atc ggt ggc tac gac ctt gat ttc ggg atc gcc gcc 528
Leu Val Ala Lys Ile Gly Gly Tyr Asp Leu Asp Phe Gly Ile Ala Ala
165 170 175

gac cag gaa ttc ata ttg ccg gcc gcg ctg gta tgc gag ccg gtc acg 576
Asp Gln Glu Phe Ile Leu Arg Ala Ala Leu Val Cys Glu Pro Val Thr
180 185 190

att ccg tgt gtg ctg tgc gag ttc gac acc acg ggc gtc ggc tcg cac 624
Ile Arg Cys Val Leu Cys Glu Phe Asp Thr Thr Gly Val Gly Ser His
195 200 205

ccg gaa cca agc gcg gtc ttc ggt gat ctg cgc cgc atg ggc gac ctt 672
Arg Glu Pro Ser Ala Val Phe Gly Asp Leu Arg Arg Met Gly Asp Leu
210 215 220

cat cgc cgc tac ccg ttc ggg gga agg cga ata tca cat gcc tac cta 720
His Arg Arg Tyr Pro Phe Gly Gly Arg Arg Ile Ser His Ala Tyr Leu
225 230 235 240

cgc ggc ccg gag ttc tac gcc tac aac agt cga ttc tgg gaa aac gtc 768
Arg Gly Arg Glu Phe Tyr Ala Tyr Asn Ser Arg Phe Trp Glu Asn Val
245 250 255

ttc acg cga atg tcg aaa tag 789
Phe Thr Arg Met Ser Lys
260

<210> 31
<211> 262
<212> PRT
<213> Mycobacterium

<400> 31
Met Thr Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Phe Asn Asp
1 5 10 15

```

Leu Asp Gly Leu Gln Arg Thr Val Lys Ser Val Arg Ala Gln Arg Tyr  
                   20                  25                  30  
 Arg Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Asp  
                   35                  40                  45  
 Val Val Ala Tyr Leu Ser Gly Cys Glu Pro Gly Phe Ala Tyr Trp Gln  
                   50                  55                  60  
 Ser Glu Pro Asp Gly Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala  
                   65                  70                  75                  80  
 His Ala Ser Gly Asp Leu Leu Trp Phe Leu His Ser Ala Asp Arg Phe  
                   85                  90                  95  
 Ser Gly Pro Asp Val Val Ala Gln Ala Val Glu Ala Leu Ser Gly Lys  
                   100                  105                  110  
 Gly Pro Val Ser Glu Leu Trp Gly Phe Gly Met Asp Arg Leu Val Gly  
                   115                  120                  125  
 Leu Asp Arg Val Arg Gly Pro Ile Pro Phe Ser Leu Arg Lys Phe Leu  
                   130                  135                  140  
 Ala Gly Lys Gln Val Val Pro His Gln Ala Ser Phe Phe Gly Ser Ser  
                   145                  150                  155                  160  
 Leu Val Ala Lys Ile Gly Gly Tyr Asp Leu Asp Phe Gly Ile Ala Ala  
                   165                  170                  175  
 Asp Gln Glu Phe Ile Leu Arg Ala Ala Leu Val Cys Glu Pro Val Thr  
                   180                  185                  190  
 Ile Arg Cys Val Leu Cys Glu Phe Asp Thr Thr Gly Val Gly Ser His  
                   195                  200                  205  
 Arg Glu Pro Ser Ala Val Phe Gly Asp Leu Arg Arg Met Gly Asp Leu  
                   210                  215                  220  
 His Arg Arg Tyr Pro Phe Gly Gly Arg Arg Ile Ser His Ala Tyr Leu  
                   225                  230                  235                  240  
 Arg Gly Arg Glu Phe Tyr Ala Tyr Asn Ser Arg Phe Trp Glu Asn Val  
                   245                  250                  255  
 Phe Thr Arg Met Ser Lys  
                   260

<210> 32

<211> 1023

<212> DNA

<213> Mycobacterium

<220>

<221> CDS

<222> (1)..(1020)



&lt;400&gt; 32

gtg aag cga gcg ctc atc acc gga atc acc ggc cag gac ggc tcg tat	48
Met Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr	
1 5 10 15	
ctc gcc gaa ctg ctg ctg gcc aag ggg tat gag gtt cac ggg ctc atc	96
Leu Ala Glu Leu Leu Leu Ala Lys Gly Tyr Glu Val His Gly Leu Ile	
20 25 30	
cgg cgc gct tcg acg ttc aac acc tcg cgg atc gat cac ctc tac gtc	144
Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val	
35 40 45	
gac ccg cac caa ccg ggc gcg cgg ctg ttt ctg cac tat ggt gac ctg	192
Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Gly Asp Leu	
50 55 60	
atc gac gga acc ccg ttg gtg acc ctg ctg agc acc atc gaa ccc gac	240
Ile Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Thr Ile Glu Pro Asp	
65 70 75 80	
gag gtg tac aac ctg gcg gcg cag tca cac gtg cgg gtg agc ttc gac	288
Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp	
85 90 95	
gaa ccc gtg cac acc ggt gac acc acc ggc atg gga tcc atg cga ctg	336
Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Met Arg Leu	
100 105 110	
ctg gaa gcc gtt ccg ctc tct ccg gtg cac tgc cgc ttc tat cag gcg	384
Leu Glu Ala Val Arg Leu Ser Arg Val His Cys Arg Phe Tyr Gln Ala	
115 120 125	
tcc tcg tcg gag atg ttc ggc gcc tcg ccg cca ccg cag aac gag ctg	432
Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Leu	
130 135 140	
acg ccg ttc tac ccg ccg tca ccg tat ggc gcc gcc aag gtc tat tcg	480
Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Tyr Ser	
145 150 155 160	
tac tgg gcg acc cgc aat tat cgc gaa gcg tac gga ttg ttc gcc gtt	528
Tyr Trp Ala Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val	
165 170 175	
aac ggc atc ttg ttc aat cac gaa tca ccg ccg cgc ggt gag acg ttc	576
Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe	
180 185 190	
gtg acc cga aag atc acc agg gcc gtg gca cgc atc aag gcc ggt atc	624
Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Lys Ala Gly Ile	
195 200 205	
cag tcc gag gtc tat atg ggc aat ctg gat gcg gtc cgc gac tgg ggg	672
Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Val Arg Asp Trp Gly	
210 215 220	
tac gcg ccc gaa tac gtc gaa ggc atg tgg ccg atg ctg cag acc gac	720
Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Thr Asp	
225 230 235 240	

gag ccc gac gac ttc gtt ttg gcg acc ggg cgc ggt ttc acc gtg cgt 768  
 Glu Pro Asp Asp Phe Val Leu Ala Thr Gly Arg Gly Phe Thr Val Arg  
 245 250 255

gag ttc gcg cgg gcc gcg ttc gag cat gcc ggt ttg gac tgg cag cag 816  
 Glu Phe Ala Arg Ala Ala Phe Glu His Ala Gly Leu Asp Trp Gln Gln  
 260 265 270

tac gtg aaa ttc gac caa cgc tat ctg cgg ccc acc gag gtg gat tcg 864  
 Tyr Val Lys Phe Asp Gln Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285

ctg atc ggc gac gcg acc aag gct gcc gaa ttg ctg ggc tgg agg gct 912  
 Leu Ile Gly Asp Ala Thr Lys Ala Ala Glu Leu Leu Gly Trp Arg Ala  
 290 295 300

tcg gtg cac act gac gag ttg gct cgg atc atg gtc gac gcg gac atg 960  
 Ser Val His Thr Asp Glu Leu Ala Arg Ile Met Val Asp Ala Asp Met  
 305 310 315 320

gcg gcg ctg gag tgc gaa ggc aag ccg tgg atc gac aag ccg atg atc 1008  
 Ala Ala Leu Glu Cys Glu Gly Lys Pro Trp Ile Asp Lys Pro Met Ile  
 325 330 335

gcc ggc cgg aca tga 1023  
 Ala Gly Arg Thr  
 340

<210> 33  
 <211> 340  
 <212> PRT  
 <213> Mycobacterium

<400> 33  
 Met Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr  
 1 5 10 15

Leu Ala Glu Leu Leu Leu Ala Lys Gly Tyr Glu Val His Gly Leu Ile  
 20 25 30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val  
 35 40 45

Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Gly Asp Leu  
 50 55 60

Ile Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Thr Ile Glu Pro Asp  
 65 70 75 80

Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp  
 85 90 95

Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Met Arg Leu  
 100 105 110

Leu Glu Ala Val Arg Leu Ser Arg Val His Cys Arg Phe Tyr Gln Ala  
 115 120 125

Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Leu  
 130 135 140

Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Tyr Ser  
 145 150 155 160  
 Tyr Trp Ala Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val  
 165 170 175  
 Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe  
 180 185 190  
 Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Lys Ala Gly Ile  
 195 200 205  
 Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Val Arg Asp Trp Gly  
 210 215 220  
 Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Thr Asp  
 225 230 235 240  
 Glu Pro Asp Asp Phe Val Leu Ala Thr Gly Arg Gly Phe Thr Val Arg  
 245 250 255  
 Glu Phe Ala Arg Ala Ala Phe Glu His Ala Gly Leu Asp Trp Gln Gln  
 260 265 270  
 Tyr Val Lys Phe Asp Gln Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser  
 275 280 285  
 Leu Ile Gly Asp Ala Thr Lys Ala Ala Glu Leu Leu Gly Trp Arg Ala  
 290 295 300  
 Ser Val His Thr Asp Glu Leu Ala Arg Ile Met Val Asp Ala Asp Met  
 305 310 315 320  
 Ala Ala Leu Glu Cys Glu Gly Lys Pro Trp Ile Asp Lys Pro Met Ile  
 325 330 335  
 Ala Gly Arg Thr  
 340

<210> 34

<211> 732

<212> DNA

<213> Mycobacterium

<220>

<221> CDS

<222> (1)..(729)

<400> 34

atg agg ctg gcc cgt cgc gct cgg aac atc ttg cgt cgc aac ggc atc 48  
 Met Arg Leu Ala Arg Arg Ala Arg Asn Ile Leu Arg Arg Asn Gly Ile  
 1 5 10 15  
 gag gtg tcg cgc tac ttt gcc gaa ctg gac tgg gaa cgc aat ttc ttg 96  
 Glu Val Ser Arg Tyr Phe Ala Glu Leu Asp Trp Glu Arg Asn Phe Leu  
 20 25 30

cgc caa ctg caa tcg cat cgg gtc agt gcc gtg ctc gat gtc ggg gcc	144
Arg Gln Leu Gln Ser His Arg Val Ser Ala Val Leu Asp Val Gly Ala	
35 40 45	
aat tcg ggg cag tac gcc agg ggt ctg cgc ggc gcg ggc ttc gcg ggc	192
Asn Ser Gly Gln Tyr Ala Arg Gly Leu Arg Gly Ala Gly Phe Ala Gly	
50 55 60	
cgc atc gtc tcg ttc gag ccg ctg ccc ggg ccc ttt gcc gtc ttg cag	240
Arg Ile Val Ser Phe Glu Pro Leu Pro Gly Pro Phe Ala Val Leu Gln	
65 70 75 80	
cgc agc gcc tcc acg gac ccg ttg tgg gaa tgc cgg cgc tgt gcg ctg	288
Arg Ser Ala Ser Thr Asp Pro Leu Trp Glu Cys Arg Arg Cys Ala Leu	
85 90 95	
ggc gat gtc gat gga acc atc tcg atc aac gtc gcc ggc aac gag ggc	336
Gly Asp Val Asp Gly Thr Ile Ser Ile Asn Val Ala Gly Asn Glu Gly	
100 105 110	
gcc agc agt tcc gtc ttg ccg atg ttg aaa cga cat cag gac gcc ttt	384
Ala Ser Ser Ser Val Leu Pro Met Leu Lys Arg His Gln Asp Ala Phe	
115 120 125	
cca cca gcc aac tac gtg ggc gcc caa cgg gtg ccg ata cat cga ctc	432
Pro Pro Ala Asn Tyr Val Gly Ala Gln Arg Val Pro Ile His Arg Leu	
130 135 140	
gat tcc gtg gct gca gac gtt ctg cgg ccc aac gat att gcg ttc ttg	480
Asp Ser Val Ala Ala Asp Val Leu Arg Pro Asn Asp Ile Ala Phe Leu	
145 150 155 160	
aag atc gac gtt caa gga ttc gag aag cag gtg atc gcg ggt ggc gat	528
Lys Ile Asp Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Gly Asp	
165 170 175	
tca acg gtg cac gac cga tgc gtc ggc atg cag ctc gag ctg tct ttc	576
Ser Thr Val His Asp Arg Cys Val Gly Met Gln Leu Glu Leu Ser Phe	
180 185 190	
cag ccg ttg tac gag ggt ggc atg ctc atc cgc gag gcg ctc gat ctc	624
Gln Pro Leu Tyr Glu Gly Gly Met Leu Ile Arg Glu Ala Leu Asp Leu	
195 200 205	
gtg gat tcg ttg ggc ttt acg ctc tcg gga ttg caa ccc ggt ttc acc	672
Val Asp Ser Leu Gly Phe Thr Leu Ser Gly Leu Gln Pro Gly Phe Thr	
210 215 220	
gac ccc cgc aac ggt cga atg ctg cag gcc gat ggc atc ttc ttc cgg	720
Asp Pro Arg Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg	
225 230 235 240	
ggc agc gat tga	732
Gly Ser Asp	

&lt;210&gt; 35

&lt;211&gt; 243

&lt;212&gt; PRT

&lt;213&gt; Mycobacterium

&lt;400&gt; 35

Met Arg Leu Ala Arg Arg Ala Arg Asn Ile Leu Arg Arg Asn Gly Ile  
 1 5 10 15  
 Glu Val Ser Arg Tyr Phe Ala Glu Leu Asp Trp Glu Arg Asn Phe Leu  
 20 25 30  
 Arg Gln Leu Gln Ser His Arg Val Ser Ala Val Leu Asp Val Gly Ala  
 35 40 45  
 Asn Ser Gly Gln Tyr Ala Arg Gly Leu Arg Gly Ala Gly Phe Ala Gly  
 50 55 60  
 Arg Ile Val Ser Phe Glu Pro Leu Pro Gly Pro Phe Ala Val Leu Gln  
 65 70 75 80  
 Arg Ser Ala Ser Thr Asp Pro Leu Trp Glu Cys Arg Arg Cys Ala Leu  
 85 90 95  
 Gly Asp Val Asp Gly Thr Ile Ser Ile Asn Val Ala Gly Asn Glu Gly  
 100 105 110  
 Ala Ser Ser Ser Val Leu Pro Met Leu Lys Arg His Gln Asp Ala Phe  
 115 120 125  
 Pro Pro Ala Asn Tyr Val Gly Ala Gln Arg Val Pro Ile His Arg Leu  
 130 135 140  
 Asp Ser Val Ala Ala Asp Val Leu Arg Pro Asn Asp Ile Ala Phe Leu  
 145 150 155 160  
 Lys Ile Asp Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Gly Asp  
 165 170 175  
 Ser Thr Val His Asp Arg Cys Val Gly Met Gln Leu Glu Leu Ser Phe  
 180 185 190  
 Gln Pro Leu Tyr Glu Gly Gly Met Leu Ile Arg Glu Ala Leu Asp Leu  
 195 200 205  
 Val Asp Ser Leu Gly Phe Thr Leu Ser Gly Leu Gln Pro Gly Phe Thr  
 210 215 220  
 Asp Pro Arg Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg  
 225 230 235 240  
 Gly Ser Asp

&lt;210&gt; 36

&lt;211&gt; 732

&lt;212&gt; DNA

&lt;213&gt; Mycobacterium

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(729)

<400> 36  
 gtg aaa tcg ttg aaa ctc gct cgt ttc atc gcg cgt agc gcc gcc ttc 48  
 Met Lys Ser Leu Lys Leu Ala Arg Phe Ile Ala Arg Ser Ala Ala Phe  
 1 5 10 15

gag gtt tcg cgc cgc tat tct gag cga gac ctg aag cac cag ttt gtg 96  
 Glu Val Ser Arg Arg Tyr Ser Glu Arg Asp Leu Lys His Gln Phe Val  
 20 25 30

aag caa ctc aaa tcg cgt cgg gta gat gtc gtt ttc gat gtc ggc gcc 144  
 Lys Gln Leu Lys Ser Arg Arg Val Asp Val Val Phe Asp Val Gly Ala  
 35 40 45

aac tca gga caa tac gcc gcc ggc ctc cgc cga gca gca tat aag ggc 192  
 Asn Ser Gly Gln Tyr Ala Ala Gly Leu Arg Arg Ala Ala Tyr Lys Gly  
 50 55 60

cgc att gtc tcg ttc gaa ccg cta tcc gga ccg ttt acg atc ttg gaa 240  
 Arg Ile Val Ser Phe Glu Pro Leu Ser Gly Pro Phe Thr Ile Leu Glu  
 65 70 75 80

agc aaa gcg tca acg gat cca ctt tgg gat tgc cgg cag cat gcg ttg 288  
 Ser Lys Ala Ser Thr Asp Pro Leu Trp Asp Cys Arg Gln His Ala Leu  
 85 90 95

ggc gat tct gat gga acg gtt acg atc aat atc gca gga aac gcc ggt 336  
 Gly Asp Ser Asp Gly Thr Val Thr Ile Asn Ile Ala Gly Asn Ala Gly  
 100 105 110

cag agc agt tcc gtc ttg ccc atg ctg aaa agt cat cag aac gct ttt 384  
 Gln Ser Ser Ser Val Leu Pro Met Leu Lys Ser His Gln Asn Ala Phe  
 115 120 125

ccc ccg gca aac tat gtc ggt acc caa gag gcg tcc ata cat cga ctt 432  
 Pro Pro Ala Asn Tyr Val Gly Thr Gln Glu Ala Ser Ile His Arg Leu  
 130 135 140

gat tcc gtg gcg cca gaa ttt cta ggc atg aac ggt gtc gct ttt ctc 480  
 Asp Ser Val Ala Pro Glu Phe Leu Gly Met Asn Gly Val Ala Phe Leu  
 145 150 155 160

aag gtc gac gtt caa ggc ttt gaa aag cag gtg ctc gcc ggg ggc aaa 528  
 Lys Val Asp Val Gln Gly Phe Glu Lys Gln Val Leu Ala Gly Gly Lys  
 165 170 175

tca acc ata gat gac cat tgc gtc ggc atg caa ctc gaa ctg tcc ttc 576  
 Ser Thr Ile Asp Asp His Cys Val Gly Met Gln Leu Glu Leu Ser Phe  
 180 185 190

ctg ccg ttg tac gaa ggt ggc atg ctc att cct gaa gcc ctc gat ctc 624  
 Leu Pro Leu Tyr Glu Gly Gly Met Leu Ile Pro Glu Ala Leu Asp Leu  
 195 200 205

gtg tat tcc ttg ggc ttc acg ttg acg gga ttg ctg cct tgt ttc att 672  
 Val Tyr Ser Leu Gly Phe Thr Leu Thr Gly Leu Leu Pro Cys Phe Ile  
 210 215 220

gat gca aat aat ggt cga atg ttg cag gcc gac ggc atc ttt ttc cgc 720  
 Asp Ala Asn Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg  
 225 230 235 240

gag gac gat tga  
Glu Asp Asp

<210> 37  
<211> 243  
<212> PRT  
<213> Mycobacterium

<400> 37  
Met Lys Ser Leu Lys Leu Ala Arg Phe Ile Ala Arg Ser Ala Ala Phe  
1 5 10 15  
Glu Val Ser Arg Arg Tyr Ser Glu Arg Asp Leu Lys His Gln Phe Val  
20 25 30  
Lys Gln Leu Lys Ser Arg Arg Val Asp Val Val Phe Asp Val Gly Ala  
35 40 45  
Asn Ser Gly Gln Tyr Ala Ala Gly Leu Arg Arg Ala Ala Tyr Lys Gly  
50 55 60  
Arg Ile Val Ser Phe Glu Pro Leu Ser Gly Pro Phe Thr Ile Leu Glu  
65 70 75 80  
Ser Lys Ala Ser Thr Asp Pro Leu Trp Asp Cys Arg Gln His Ala Leu  
85 90 95  
Gly Asp Ser Asp Gly Thr Val Thr Ile Asn Ile Ala Gly Asn Ala Gly  
100 105 110  
Gln Ser Ser Ser Val Leu Pro Met Leu Lys Ser His Gln Asn Ala Phe  
115 120 125  
Pro Pro Ala Asn Tyr Val Gly Thr Gln Glu Ala Ser Ile His Arg Leu  
130 135 140  
Asp Ser Val Ala Pro Glu Phe Leu Gly Met Asn Gly Val Ala Phe Leu  
145 150 155 160  
Lys Val Asp Val Gln Gly Phe Glu Lys Gln Val Leu Ala Gly Gly Lys  
165 170 175  
Ser Thr Ile Asp Asp His Cys Val Gly Met Gln Leu Glu Leu Ser Phe  
180 185 190  
Leu Pro Leu Tyr Glu Gly Gly Met Leu Ile Pro Glu Ala Leu Asp Leu  
195 200 205  
Val Tyr Ser Leu Gly Phe Thr Leu Thr Gly Leu Leu Pro Cys Phe Ile  
210 215 220  
Asp Ala Asn Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg  
225 230 235 240  
Glu Asp Asp

<210> 38  
<211> 828

&lt;212&gt; DNA

&lt;213&gt; Mycobacterium

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(825)

&lt;400&gt; 38

```

atg gtg cag acg aaa cga tac gcc ggc ttg acc gca gct aac aca aag 48
Met Val Gln Thr Lys Arg Tyr Ala Gly Leu Thr Ala Ala Asn Thr Lys
  1             5             10             15

aaa gtc gcc atg gcc gca cca atg ttt tgc atc atc atc ccc acc ttg 96
Lys Val Ala Met Ala Ala Pro Met Phe Ser Ile Ile Ile Pro Thr Leu
             20             25             30

aac gtg gct gcg gta ttg cct gcc tgc ctc gac agc atc gcc cgt cag 144
Asn Val Ala Ala Val Leu Pro Ala Cys Leu Asp Ser Ile Ala Arg Gln
             35             40             45

acc tgc ggt gac ttc gag ctg gta ctg gtc gac ggc ggc tgc acg gac 192
Thr Cys Gly Asp Phe Glu Val Leu Val Asp Gly Gly Ser Thr Asp
             50             55             60

gaa acc ctc gac atc gcc aac att ttc gcc ccc aac ctc ggc gag cgg 240
Glu Thr Leu Asp Ile Ala Asn Ile Phe Ala Pro Asn Leu Gly Glu Arg
             65             70             75             80

ttg atc att cat cgc gac acc gac cag ggc gtc tac gac gcc atg aac 288
Leu Ile Ile His Arg Asp Thr Asp Gln Gly Val Tyr Asp Ala Met Asn
             85             90             95

cgc ggc gtg gac ctg gcc acc gga acg tgg ttg ctc ttt ctg ggc gcg 336
Arg Gly Val Asp Leu Ala Thr Gly Thr Trp Leu Leu Phe Leu Gly Ala
             100             105             110

gac gac agc ctg tac gag gct gac acc ctg gcg cgg gtg gcc gcc ttc 384
Asp Asp Ser Leu Tyr Glu Ala Asp Thr Leu Ala Arg Val Ala Ala Phe
             115             120             125

att ggc gaa cac gag ccc agc gat ctg gta tat ggc gac gtg atc atg 432
Ile Gly Glu His Glu Pro Ser Asp Leu Val Tyr Gly Asp Val Ile Met
             130             135             140

cgc tca acc aat ttc cgc tgg ggt ggc gcc ttc gac ctc gac cgt ctg 480
Arg Ser Thr Asn Phe Arg Trp Gly Gly Ala Phe Asp Leu Asp Arg Leu
             145             150             155             160

ttg ttc aag cgc aac atc tgc cat cag gcg atc ttc tac cgc cgc gga 528
Leu Phe Lys Arg Asn Ile Cys His Gln Ala Ile Phe Tyr Arg Arg Gly
             165             170             175

ctc ttc ggc acc atc ggt ccc tac aac ctc cgc tac cgg gtc ctg gcc 576
Leu Phe Gly Thr Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Leu Ala
             180             185             190

gac tgg gac ttc aat att cgc tgc ttt tcc aac cca gcg ctc gtc acc 624
Asp Trp Asp Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Val Thr
             195             200             205

```



cgc tac atg cac gtg gtc gtt gca agc tac aac gaa ttc ggc ggg ctc 672  
 Arg Tyr Met His Val Val Val Ala Ser Tyr Asn Glu Phe Gly Gly Leu  
 210 215 220  
 agc aat acg atc gtc gac aag gag ttt ttg aag cgg ctg ccg atg tcc 720  
 Ser Asn Thr Ile Val Asp Lys Glu Phe Leu Lys Arg Leu Pro Met Ser  
 225 230 235 240  
 acg aga ctc ggc ata agg ctg gtc ata gtt ctg gtg cgc agg tgg cca 768  
 Thr Arg Leu Gly Ile Arg Leu Val Ile Val Leu Val Arg Arg Trp Pro  
 245 250 255  
 aag gtg atc agc agg gcc atg gta atg cgc acc gtc att tct tgg cgg 816  
 Lys Val Ile Ser Arg Ala Met Val Met Arg Thr Val Ile Ser Trp Arg  
 260 265 270  
 cgc cga cgt tag 828  
 Arg Arg Arg  
 275

<210> 39  
 <211> 275  
 <212> PRT  
 <213> Mycobacterium

<400> 39  
 Met Val Gln Thr Lys Arg Tyr Ala Gly Leu Thr Ala Ala Asn Thr Lys  
 1 5 10 15  
 Lys Val Ala Met Ala Ala Pro Met Phe Ser Ile Ile Ile Pro Thr Leu  
 20 25 30  
 Asn Val Ala Ala Val Leu Pro Ala Cys Leu Asp Ser Ile Ala Arg Gln  
 35 40 45  
 Thr Cys Gly Asp Phe Glu Leu Val Leu Val Asp Gly Gly Ser Thr Asp  
 50 55 60  
 Glu Thr Leu Asp Ile Ala Asn Ile Phe Ala Pro Asn Leu Gly Glu Arg  
 65 70 75 80  
 Leu Ile Ile His Arg Asp Thr Asp Gln Gly Val Tyr Asp Ala Met Asn  
 85 90 95  
 Arg Gly Val Asp Leu Ala Thr Gly Thr Trp Leu Leu Phe Leu Gly Ala  
 100 105 110  
 Asp Asp Ser Leu Tyr Glu Ala Asp Thr Leu Ala Arg Val Ala Ala Phe  
 115 120 125  
 Ile Gly Glu His Glu Pro Ser Asp Leu Val Tyr Gly Asp Val Ile Met  
 130 135 140  
 Arg Ser Thr Asn Phe Arg Trp Gly Gly Ala Phe Asp Leu Asp Arg Leu  
 145 150 155 160  
 Leu Phe Lys Arg Asn Ile Cys His Gln Ala Ile Phe Tyr Arg Arg Gly  
 165 170 175

Leu Phe Gly Thr Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Leu Ala  
 180 185 190  
 Asp Trp Asp Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Val Thr  
 195 200 205  
 Arg Tyr Met His Val Val Val Ala Ser Tyr Asn Glu Phe Gly Gly Leu  
 210 215 220  
 Ser Asn Thr Ile Val Asp Lys Glu Phe Leu Lys Arg Leu Pro Met Ser  
 225 230 235 240  
 Thr Arg Leu Gly Ile Arg Leu Val Ile Val Leu Val Arg Arg Trp Pro  
 245 250 255  
 Lys Val Ile Ser Arg Ala Met Val Met Arg Thr Val Ile Ser Trp Arg  
 260 265 270  
 Arg Arg Arg  
 275

<210> 40  
 <211> 24  
 <212> DNA  
 <213> Mycobacterium

<400> 40  
 gatgccgtga ggaggtaaag ctgc

24

<210> 41  
 <211> 24  
 <212> DNA  
 <213> Mycobacterium

<400> 41  
 gatacggctc ttgaatcctg cacg

24